Are you asking me or telling me: A Study of L2 Speakers' Perception of Two Emotional and Pragmatic Cues in Intonationally Complex Sentences in Mandarin Chinese

by

Yibo Liu

(Under the Direction of Linda Harklau)

Abstract

Mandarin Chinese, like other tonal languages, makes use of modulation in Fo for both marking lexical items and pragmatic purposes. This makes it challenging for L2 speakers to acquire an accurate perception of prosody and identify emotional and pragmatic intents. This thesis presents a study on how L2 Mandarin speakers perceive two emotional and pragmatic cues, namely emphaticness and confidence, in intonationally complex sentences in Mandarin. Results show that compared to L2 speakers, monolingual English speakers consistently rated emphatic sentences as less emphatic, which suggests that L2 speakers were able to acquire at least some intonational knowledge of emphatic sentences in Mandarin. On the other hand, when compared to L2 speakers, native Mandarin speakers consistently rated questions as sounding less confident than L2 speakers did. This suggests that the prosody of questions in Mandarin is a challenging component to acquire. Further research on prosody in Mandarin and cross-linguistic comparisons between prosodies of different languages is needed to further address L2 Mandarin learners' difficulty in acquiring pragmatic prosody in the language.

Index words: [Second Language Acquisition, Intonation, Prosody, L2-Mandarin, Pragmatics]

Are you asking me or telling me: A Study of L2 Speakers' Perception of Two Emotional and Pragmatic Cues in Intonationally Complex Sentences in Mandarin Chinese

by

Yibo Liu

B.Ed. Corban University, 2022

B.Ed. Universitas Pelita Harapan, Indonesia, 2022

A Thesis Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Master of Arts

Athens, Georgia

2025

Are you asking me or telling me: A Study of L2 Speakers' Perception of Two Emotional and Pragmatic Cues in Intonationally Complex Sentences in Mandarin Chinese

by

Yibo Liu

Major Professor: Linda Harklau

Committee: Pillar Chamorro

Jon Forrest

Electronic Version Approved:

Ron Walcott

Dean of the Graduate School

The University of Georgia

August 2025

Dedication

To Him Who grants me strength, courage, wisdom, and intellect

To my dearest Mom, Dad, and Tiffany, who are my biggest fans

To Ms. Rena Weng, whose teaching and love for language cultivated the spark that led me down this path

Acknowledgments

I'd like to begin this acknowledgements section by first stating that this project, which has been the biggest undertaking in my life by far, could not have possibly come into fruition, had it not been for the countless individuals and kind souls who have assisted me in all kinds of ways, shapes, and forms throughout the entire process. This very short acknowledgement is in no way an exhaustive list of all those who have played a role, no matter how big or small, in helping me finish this endeavor of mine, and I owe every one of them my sincerest thanks and gratitude.

First, I'd like to thank my dearest Mom and Dad for making possible my education. Mom and Dad, thank you two so much for having raised me from the little boy that I was in Xinmi to the young man that I have become today. I thank you two for all the hassles you had to go through to send me abroad for my education, and I could never thank you two enough for having supported my passion and dreams all these years.

Finishing this thesis project has been an arduous journey, one that I could not have completed myself without the help of my thesis committee members. My deepest thanks goes to my advisor, Dr. Linda Harklau, who generously accepted me as her advisee despite not having worked with me beforehand and whose expertise and mentorship helped me greatly while I was still approaching research as an inexperienced rookie. I'm also incredibly grateful for Dr. Jon Forrest, who generously offered his help for doing statistics and data modeling, and Dr. Pillar Chamorro, whose expertise in pragmatics helped me interpret my data in more meaningful ways.

In addition to my committee members, I'd like to note that this thesis project was done through the Second Language Acquisition and Cognition Lab (SLAC) at the Department of Linguistics at UGA. And I owe it every bit of my gratitude as well. I'd especially like to thank SLAC coordinator and my friend Jean Costa-Silva, who offered to take me in when I needed guidance and peer support for doing my research,

and who has offered much of his generous help and mentorship throughout the process. My thanks also goes to Caroline Schneider, who offered her help and feedback during the initial planning stage of this project, and Sarah Moore, who has helped me with recording the stimuli that were used in this project and who also reached out to me personally with her thoughts, prayers, and encouragement.

I would also like to extend my gratitude to those whose support, though not directly related to this project, has been invaluable to me on a personal level. To all my friends, peers, and the uncles and aunties at Athens Chinese Christian Church—thank you so much for your continued thoughts and prayers for both my academic endeavors and my well-being. To my friend and roommate, Bingxi, for having supported me in countless personal ways—whether it was driving me to get groceries, taking me out for Chinese food, praying for me during my darkest hours, or helping me in ways I can no longer even recall. To my friends, Miles and Mara Paca, who have welcomed me into their home time and again for breakfast and dinner, and many a time have we exchanged heartfelt conversations together. To my long time friend, Sheryl Chen, who consistently reaches out to me with her thoughts, prayers, and heartfelt blessings. To my friends over at CGF Monday morning fellowship, Annabelle and her husband Jackson Hill, who drove me to their church's weekly Saturday Shabbat dinner many times and whose friendship I will always remember, and Kelly Tims, for her continued thoughts and prayers.

Last but definitely not least, special thanks also goes to our beloved Amy Smoler in the linguistics department, who always keeps things running smoothly and who has made the lives of all graduate students there easier.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Many languages, including English, use changes in the Fo contour to signal pragmatic purposes, such as asking questions, expressing surprise, and signifying authority (Yang et al., 2021). "Fo contour" refers to the fundamental frequency, namely the vibration of a tiny piece of flesh called the vocal folds that create various phonological qualities of human voices, including but not limited to the vowel sounds one can produce, the pitch of one's voice, etc. (Zsiga, 2024). Intonation and tones are part of the phonological processes that utilize this fundamental frequency to achieve linguistic meaning.

Mandarin Chinese, like many other tonal languages, uses modulations in the Fo contour for the realization of lexical tones. Chao, in one of his seminal works on spoken Mandarin, described Mandarin as having four or five lexical tones, depending on whether the neutral tone is counted as one (Y. R. Chao, 1948). Morphophonological realization of pragmatic cues in Mandarin often includes grammatical markers, which are otherwise more commonly known as particles (Y.-R. Chao, 1933). A question, for example, would end with the particle "ma (吗)". A surprise question, on the other hand, would often end with the particle "a (啊)". However, while particles are predominantly used in mundane discourse for realizing prosodic cues, modulations in the Fo contour are another morphophonological tool also utilized in Mandarin for the realization of prosodic cues (Yang et al., 2021). It could reasonably be predicted, therefore, that such parallel use of modulations in the Fo contour presents cognitive challenges to listeners, particularly on the psychoacoustic level.

This study, therefore, sought to answer the question of how intonational patterns within tonal languages were perceived by L2 speakers, specifically L2 speakers of Mandarin who speak English as their L1. The results for L2 speakers were contrasted against those of native speakers of Mandarin and naive speakers of English who do not possess any knowledge of tonal languages such as Mandarin. Specifically, this study aimed to discover how emotional and pragmatic cues are perceived, and whether they are perceived similarly or differently, by L2 speakers, compared to native Mandarin speakers and naive English speakers. By looking at how they perceive emotional and pragmatic cues, we may be able to identify which pragmatic cues are more or less accessible to L2 learners. This would help us understand not only how prosody works differently with respect to pragmatics in both languages, but also the general patterns for how prosody works in languages and how it is acquired in language learning.

1.2 Tonality, Intonation, Prosody and the Suprasegmentals

Human beings possess the ability to extract information from physical stimuli and abstract such information into mental representations. This remarkable ability to process and interpret the physical as meaningful mental representations is what makes communication possible. On a neurocognitive level, it would seem that the human brain has the capability to extract certain acoustic cues from the raw, physical sounds that it receives. According to our current understanding of the neurocognitive basis for processing speech sounds, the human brain processes physical, acoustic sounds through both phonetic-acoustic and phonological means (Mai et al., 2024). This parallel processing model suggests that the human brain is capable of both processing the raw, physical sound and interpreting it using abstract notions, such as the idea of a segment, and by extension, a phoneme. After the raw, physical sound signals are processed and extracted into more abstract representations (which linguists refer to as phonemes), these phonemic representations are then strung together to form meaningful units such as morphemes, lexemes, phrases, and sentences.

What is so interesting about speech processing and perception is that while we do have solid neurocognitive evidence for the processing of the raw, acoustic signals into the abstract representations that we harbor in our minds, a more fine-grained explanation for how this works in detail is still a work in progress. In particular, we still lack a thorough explanation for how the brain processes non-spectral acous-

tic cues—such as modulations in the Fo contour or intensity—and how these perceptual mechanisms contribute to the interpretation that the brain generates.

Linguistics in the 20th century had unfortunately neglected much of the suprasegmentals in their development of phonological theories and had also attempted to incorporate them into the more traditional segmental framework. It was not until fairly recently that linguists began to realize the importance of developing a more comprehensive framework that grants the suprasegmental a more theoretically thorough and accurate treatment. One thing in particular about the suprasegmentals which merits such treatment has to do with how the suprasegmentals are both articulatorily and conceptually different from the segmentals. Articulatorily, the features that linguists refer to as the segmentals, namely vowel formants, place of articulation, manner of articulation, etc., are produced in the supralaryngeal portion of the speech organ apparatus (Abercrombie, 1967; Lieberman & Blumstein, 1988). Features such as intonation (pitch), intensity, voice quality, etc., which collectively contribute to prosodic patterns, on the other hand, are produced in the sublaryngeal portion of the speech organ apparatus. Conceptually, features that are produced by the supralaryngeal portion predominantly contribute to the lexical and morphosyntactic interpretation of an utterance, whereas sublaryngeal phonological components such as intonation, intensity, voice quality, etc., contribute to the interpretation of pragmatic intent, such as inquiry, statement, and emphasis (Gussenhoven, 2004). Additionally, it has also been argued that paralinguistic functions such as the expression of emotions are also encoded in the aforementioned sublaryngeal acoustic features, which collectively form the phonological component that is more commonly known by linguists as prosody (Ladd, 2008).

Aside from marking pragmatic intent and paralinguistic information, suprasegmental features are also used for marking important lexical information in languages that do have such a parameter. For some languages, the use of such a parameter is manifested in the form of certain syllables bearing more prominence by having a higher pitch. This phenomenon is often referred to as pitch accent (Zsiga, 2024). The other more prominent manner in which this parameter is manifested is what phonologists refer to as lexical tones. Lexical tones are Fo modulations at the word level that contribute to its lexical meaning. Mandarin Chinese, for example, is one such language that has four [1] lexical tones. The tones in Mandarin are integral for interpreting lexical information from utterances. Pronouncing the syllable ma˥ as ma˧˥, for

[1] or 5 depending on analysis. The neutral tone is often described as an additional tone, but some phonologists argue that it is simply a lack of tone rather than a distinct tonal category

example, will result in a different lexical item, as the former means "mother", and the latter means "horse". The fact that Mandarin uses Fo for marking lexical information would necessarily warrant a discussion about its implications for how Mandarin manages pragmatics and paralinguistic information, as such information is shown to be cross-linguistically encoded in modulations of the Fo contour (Romero-Trillo & Sadeghi, 2023).

It is reasonable to postulate that because Mandarin Chinese utilizes Fo for both lexical and pragmatic information, there is going to be an interaction between the two. In particular, there is evidence suggesting that Mandarin speakers may utilize an additional part of their brain when processing Fo contours. Chien et al. (2020), based on their research in fMRI imaging, claim that when exposed to monosyllabic input with tones, Mandarin speakers demonstrated additional neural activity in the bilateral temporo-parietal regions, which are typically associated with semantic processing, and in the subcortical areas, which are associated with processing of tones and intonation (Chien et al., 2020). Prior research has demonstrated that this interaction could manifest in the form of additional cognitive load on Mandarin speakers. Specifically, Mandarin speakers have been shown to have longer reaction times and are less accurate at identifying question intonational contours when compared to bilingual Uyghur speakers (Liang & van Heuven, 2007). Additionally, a study has also demonstrated that Mandarin native speakers are less attentive to prosodic patterns when asked to identify paralinguistic cues, compared to non-tonal language speakers[2] (Zhu, 2013).

Based on the above discussion, it could be reasonably concluded that lexical tones in Mandarin do have a bearing on how speakers perceive intonational and prosodic patterns, particularly those having to do with pragmatic and paralinguistic functions. Several research questions could therefore be raised, based on our current understanding of prosody, intonation, and lexical tones. Before delving into core research questions, it is necessary to establish a common framework about tones, intonation, and prosody, within which hypotheses can be constructed and tested. To achieve that, in Chapter 2 I provide a comprehensive, albeit not entirely exhaustive, review of relevant literature and studies that have been done in the past on tones, intonation, and prosody in Mandarin. Specifically, I focus on studies done to investigate the effects of lexical tones on the perception of intonation and prosody in Mandarin. In Chapter 3 I articulate

[2]Dutch speakers and Dutch L2 Mandarin speakers

research questions and detail the study methodology. In Chapter 4 I present and discuss study results. In Chapter 5 I provide implications for further research.

Chapter 2

Background

2.1 Tones, Prosody, and Intonation

2.1.1 Tones

Historically, tones, intonation, and prosody have been treated by phonologists as three components that operate on the same phonological hierarchy, which many phonologists refer to as the suprasegmentals. As a result, tones, intonation, and prosody, have historically been treated by phonologists as different processes which make use of the same phonetic correlates (i.e., modulation in the Fo contour). Mandarin, being the most extensively studied tonal language, provides us with tonal paradigms that give us insights into how tone works in language. Chao (Y. R. Chao, 1968, p.39), when discussing the interaction between tones, intonation, and prosody, used the following metaphor to describe it:

> The best answer is to compare syllabic tone and sentence intonation with small ripples riding on large waves (though occasionally the ripples may be "larger" than the waves). The actual result is an algebraic sum of the two kinds of waves.

and:

> Stretching the [elastic transparent] sheet or letting it shrink will vary the pitch range of tones; doing the same horizontally will make the syllables longer or shorter.

(Y. R. Chao, 1968, p.26)

Chao's argument proposes that lexical tones act like perturbations in the overall Fo contour of the speaker. As a result, it implies that the locutionary/pragmatic intent of the speaker would predetermine the overall shape of the Fo contour, and that lexical tones, which are phonologically realized as perturbations on the predetermined Fo contour, are determined and realized after prosody and intonation.

This kind of model that treats intonation as a separate phonological component that is independent from lexical tones has been widely recognized in earlier studies of Mandarin intonation. In these earlier studies, intonational cues (such as those of a question, statement, etc.) are treated as either a rising or falling curve along the Fo contour (Gårding, 1987; Shen, 1990). In more recent studies, this idea is still largely being embraced (Fujisaki et al., 2005; Ni & Hirose, 2006; Prom-On et al., 2009; Yuan et al., 2002). For example, Yuan's model of intonational patterns on the Fo contour uses Stem-ML (Soft Template Mark-up Language). Yuan and colleagues' study, based on their modeling of the Fo contour, concludes that the manifestation of a question intonation consists of two mechanisms. These two mechanisms include an overall higher Fo curve on the phrase level, and a sharper rise at sentence final syllables, particularly the boundary contour of sentence endings (Yuan et al., 2002).

It should be noted, though, that it is still a debate within Sinitic linguistics whether tones and intonations should be treated as the same or different phonological categories. In addition, it is even more debated whether intonation patterns are encoded globally along the contour of a whole utterance, or simply at the boundary contour of sentence endings.

2.1.2 Prosody and Intonation

Prosody and intonation are generally what linguists use to describe the collective features that go beyond the segmental level. These are often called suprasegmentals, and they include features such as intensity (which is often manifested in the form of duration and amplitude), and pitch (which is manifested in the form of changes in the fundamental frequency). Any utterance is going to inevitably encode some type of prosody, no matter what language one speaks. The term prosody usually encodes more information than the term intonation does, as the former involves further suprasegmental information such as intensity, amplitude, etc. Intonation, therefore, is a subcomponent of prosody. One of the most important functions that intonation serves is encoding discourse information that is otherwise not readily apparent

(Cutler et al., 1997). One of the many ways in which intonational prosody is used to encode discourse information is that it expresses focus. This particular speech act is achieved syntactically through word order in other languages like Catalan or Italian, but is nonetheless realized through intonational patterns in English(Ladd, 2008). In short, prosody is what speakers of many different languages use to express discourse and pragmatic information. And because intonation is a type of prosody, any research that involves pragmatics and prosody is inevitably going to involve intonation as well.

Similar to how they work in other languages, prosody and intonation are also extensively used for expressing linguistic information that is not encoded morphosyntactically. In Mandarin Chinese, pragmatic and discourse information is often encoded using a special class of words called particles. These particles are usually devoid of concrete, semantic meanings themselves and are only ever used in sentences and set phrases to mark discourse or pragmatic intent (Y. R. Chao, 1948). These particles, while technically required, are optional in the context of sentences where such discourse or pragmatic intent is expressed and encoded in the form of intonations, which, in turn, are going to correspond to modulations and disturbances on the Fo contour of a particular utterance. Specifically, a question that is usually realized in the following form with a particle:

1. 他今天敲钟吗

   $t^ha˥$ $tɕin˥$ $t^hiən˥$ $tɕ^hao˥$ $dʒoŋ˥$ $ma˥$
   He now day hit bell prt.Q

   Is he ringing the bell today?

could instead become a sentence with the following pitch contour at the end:

1. 他今天敲钟

   $t^ha˥$ $tɕin˥$ $t^hiən˥$ $tɕ^hao˥$ $dʒoŋ˥$
   He now day hit bell

   Is he ringing the bell today?

The final syllable/character, while bearing the first tone [1], is affected by the pragmatic intent of the speaker and therefore gets a rising contour isntead of a level one. This type of sentence, as the readers shall see, will

[1] This is usually a mid, level tone if it was to be described using linguistically neutral terminologies

be the central focus of this research, as they provide us valuable prototypes for testing speakers' localization and interpretation of pragmatic cues within utterances that would otherwise have been morphosyntactically identical and therefore ambiguous.

To sum up this section, tones, intonation, and prosody, are three closely related yet distinct phonological categories within the phonological hierarchy. Because lexical tones and prosodic intonation are often realized in similar phonetic forms, namely modulations on the Fo contour, they are often seen and categorized as one phonological category. And because of such, the interaction between lexical tones and intonation, particularly the intonational patterns of tonal languages, are severely understudied.

Regardless of the unfortunate status quo of the field, intonational patterns, nonetheless, are an important phonological component, carrying vital linguistic information such as discourse, pragmatic intent, and paralinguistic features such as emotion as well. As a result, they would still require further studies and investigation. Moreover, the fact that tonal language speakers use Fo contours to express both lexical and prosodic information requires further study on its effects on intonational patterns. For one thing, Mandarin speakers have been shown to have evoked both the brain region that is responsible for semantic processing and the brain region for tonal processing when exposed to stimuli, indicating that they are capable of processing both at the same time (Chien et al., 2020). Research done by Liang and van Heuven, 2007 and Zhu, 2013 suggests this simultaneous processing may in fact contribute additional processing load for the speaker, making them worse at making judgements about a speaker's pragmatic intent based on the Fo contour alone. In addition, there is still ongoing debate in the field as to whether lexical tones and intonational contours should be treated the same or different. Scholars such as Ladd, 2008, for example, view them as essentially different instantiations of the same phonological component in the prosodic hierarchy. Others suggest that they should be treated as (Fujisaki et al., 2005; Ni & Hirose, 2006; Prom-On et al., 2009; Yuan et al., 2002). Y. R. Chao, 1968, on the other hand, had repeatedly used the metaphor of intonation being "small ripples riding on large waves", therefore offering a more nuanced view on the matter. All of the above would grant impetus for researchers to conduct studies in this field.

2.2 Speech Acts, Intonation, and Prosody

As had been established previously, intonational and prosodic patterns often become the phonological correlates for encoding pragmatic, and sometimes paralinguistic information. In this section, I intend to examine how questions and statements are phonologically described and realized. Specifically, I intend to examine existing literature on the prosodic and intonational patterns of questions and statements in Mandarin and English and compare them.

2.2.1 The Pragmatics of Speech Acts

One of the central subfields of the study of pragmatics in linguistics is the study of speech acts. Speech acts, in general, could be defined as the notion that human utterances are not simply vehicles by which meanings are transferred and imparted, but rather means through which concrete actions are taken and purposes and intents achieved (Huang, 2017). Scholars like J.L. Austin, for example, have argued that historical philosophies of language have placed much of their emphasis on the study of statements, or propositions, and in doing so have failed to recognize the actual functions that everyday-language serves (Austin, 1975). This notion is quite literally reflected in the name of his seminal work on pragmatics titled "How to do things with words". Austin's seminal work was later expanded upon by scholars such as John Searle, who connected Austin's theory to sociology and linguistics. He went on to suggest that a multitude of typologies of speech acts could be derived by clustering types of felicity conditions, so that there can be only five main types of speech acts: representatives (statements and the like), directives (questions, requests, orders), commissives (threats, promises, offers), expressives (thanking, apologizing, congratulating, etc.), and declarations (like christening, declaring war, firing, etc. (Searle, 1969). There are other fine-grained details concerning speech acts that I could get into. Since this study concerns itself with the interface between the pragmatics of speech acts and prosody on the linguistics side, and also for the sake of brevity, I choose to not dabble in much detail regarding the validity of Searle's framework and instead will focus primarily on the interface between the pragmatics of speech acts and prosody.

2.2.2 The Interface between Pragmatics and Prosody

It has been well established that variation in prosody can influence the interpretation of linguistic content significantly in many languages (Ladd, 2008). Prosody, as a phonological component, is primarily encoded in the Fo contour. When it comes to studying the interface between this phonological component and pragmatics, particularly the type that has to do with speech acts, there are mainly two approaches. These two approaches are best characterized as "the continuous" and "the categorical" (Huang, 2017). The continuous approach mainly focuses on studying prosody as variations on the Fo contour in the same manner that one would study a physical continuum. The categorical approach, on the other hand, mainly focuses on studying prosody as distinct segments that correspond to particular prosodic, linguistic, and speech act phenomena. One shall see in the following sections, however, that this binary distinction between theoretical approaches to the interface between pragmatics and prosody is not enough for studies concerning such a subtopic in Mandarin.

2.2.3 Sentence-Level Prosody in Mandarin

Defining and describing the intonational and prosodic patterns of sentence-level utterances in Mandarin is challenging. The tonal nature of the language, combined with its propensity for tone sandhi [2] and the interaction between lexical tones and sentence-level intonational patterns, makes it an arduous, layered task. It is layered because several phonological components affect the observed Fo contour of an utterance. These include lexical tones, stress, and sentence-level intonational patterns based on pragmatics/syntax (Shen, 1990). And this is to say nothing of paralinguistic emotional affectations that could have an effect on the eventual Fo contour.

To examine this layered interaction, it is imperative that one start with the interaction between lexical tones and intonation. As had been mentioned previously in 2.1.2, there is currently no consensus on how the interaction between lexical tones and sentence-level intonation should be described, although there are several competing alternatives. In the broad body of literature concerning this topic, there are two views that deserve attention. Y. R. Chao, 1968 suggests that sentence-level intonational patterns act as

[2] Tone Sandhi refers to phonological changes that happen because of the interactions between tones. This could manifest in a changed tone or a changed tone with other suprasegmental changes, such as stress. Y. R. Chao, 1948 gave a fairly thorough description of some commonly observed sandhis in Mandarin in his primer, but it will not be discussed at length in this study as it is not fundamentally relevant to my research questions

perturbations on the existing pitch contour of the lexical tones [3]. In particular, in one of his earlier works, he assumes that there are two types of intonational "additions": simultaneous and successive (Y.-R. Chao, 1933, p. 131). By "simultaneous" he means that one of the following four features is simultaneously added to lexical tones to form the resultant sentence-level Fo contour:

1. general raised level of pitch
2. general lowered level of pitch
3. pitch range widened
4. pitch range narrowed

By "successive", he means that the rising and falling endings of intonation are added after the mapping of tone, and therefore the tone value as well as tone shape are changed according to the direction of the endings. This is better show in 2.1[4]:

Rising Endings	Falling Endings
55 = 56	55 = 551
35 = 36	35 = 351
214 = 216	214 = 2141
51 = 513	51 = 5121

Figure 2.1: Chao's modification of tones

Before I go any further, it should be noted that Chao used a system of numbers to describe tonal contours in both lexical tones and intonation, with number 1 referring to the lowest level tone and 5 the highest level tone, and number combinations referring to any rising/falling contours that would go in between.

This approach to the interaction between lexical tones and intonation can also be found in other scholars' works. Although Rumjancev (1978)[143] does not explicitly say that intonation is successively added

[3]See 2.1.1 where I gave the direct quote from his Mandarin Primer

[4]The formalism that Chao uses is one he himself invented.Instead of using shapes and lines, he opted to use numbers, with 1 representing the lowest pitch and higher numbers representing higher pitches. A 214 would, for example, be a falling-rising tone, as the pitch level drops from 2 to 1, and then rises to 4

Tones	Rising ending	Falling ending
high-rising	high-rising	mid-level
low-falling	low-level	low-falling
high-falling	high-level	high-falling
low-falling-rising	low-rising	low-falling

Figure 2.2: Chang's modification of tones

Note:It is difficult to give a more thorough, yet terse explanation for what exactly Chang was trying to argue for here. It should be understood, however, that Chang's main point was that intonational fortition/lenition of lexical tones are not strictly algebraic, which Chao was arguing for in his works

to tone, the phenomenon he calls "supplementary post-tonal stretches of phonation" could be interpreted as such. To sum it up, scholars such as Chao and Rumjancev believe that the general intonational pattern of an utterance in Mandarin consists of a two-fold process. One process is referred to as simultaneous addition. Essentially, the intonational pattern of an utterance is algebraically added (or subtracted) to each lexical tone of the entire utterance. Successive addition, on the other hand, refers to the concatenative addition that is superimposed on the terminal contour. Y.-R. Chao, 1933 assumes a terminal rise or fall as intonational cues for questions and statements in Mandarin. Therefore, a falling intonation would result in slight reduction to pre-terminal lexical tones and a noticeable drop to the terminal contour. A rising intonation would, on the other hand, result in a slight addition to pre-terminal lexical tones and a noticeable rise to the terminal contour.

There are of course, scholars who disagree with Chao with regards to his approach to the interaction between lexical tones and intonation in Mandarin, the most notable ones of which include Chang (1958) and Ho (1976). Similar to Chao, they agree with the notion of sentence-level intonational patterns being superimposed onto lexical tones. Chang, 1958, however, based on her research, proposes that at least in one variety of Mandarin[5], the superimposition of sentence-level intonation does not result in concatenative changes to all lexical tones involved. Chang was only able to notice terminal rising/falling, based on her phonetic measurements of utterances produced by native speakers of that variety. Chang's finding for what happens at the terminal contour could be best summarized by Figure 2.2. Ho, 1976, while agreeing with Chao's notion of superimposition happening at the level of all lexical tones, does not adhere to the idea of

[5]Sichuan Mandarin, to be more specific

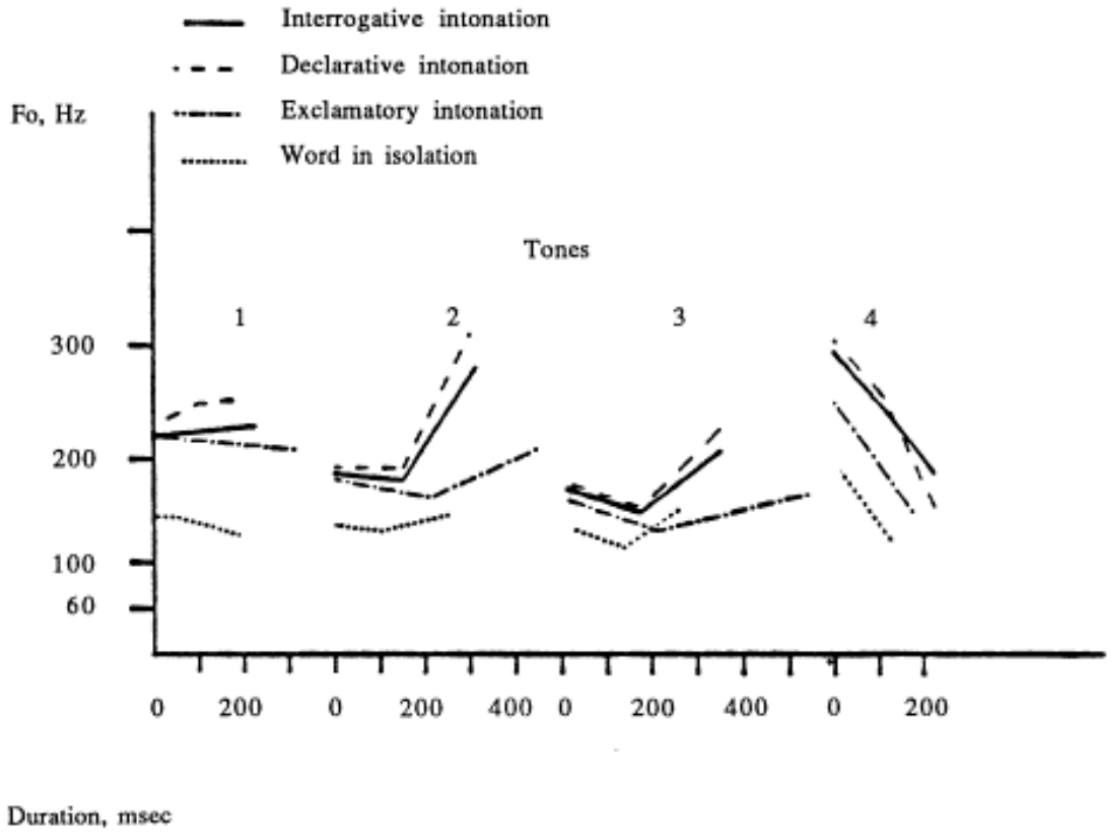

Figure 2.3: Ho's modification of tones in three intonation environments compared with each tone in isolation

concatenative addition. Her evidence lies in the fact that the basic contour of lexical tones are neutralized when the direction of the intonational contour differs from that of the lexical tones. Moreoever, when they happen to overlap, such contour features become more salient. Hence, one might argue, based on their evidence, that phonological fortition and lenition happens at the contour level as a result of the superimposition of sentence-level intonational patterns onto lexical tones. Figure 2.3 shows Ho's description of intonation-tonal interaction.

This idea of a global superimposition of intonational patterns could also be found in scholarly literature such as that of Gårding, 1987's. Gårding, 1987 proposed a framework called "the grid", which was originally used for one of his other studies conducted on intonational patterns in Swedish (Gårding, 1979).

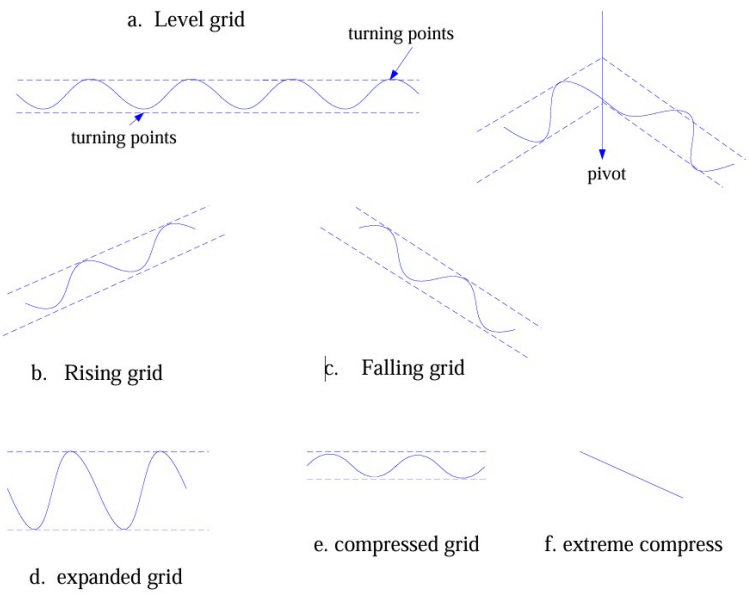

Figure 2.4: Gårding, 1993's grid framework

According to the grid model, intonation pattern is "the global melody of a phrase or sentence to which the local pitch movements are subordinated"(Gårding, 1993, p. 36). Therefore, according to Gårding, 1993, sentence-level intonational patterns could be best characterized as one that consists of two invisible lines (i.e., the grid) that demarcate the upper and lower boundaries of the pitch contour. The direction of the grid reflects the global rising/falling pattern of a sentence-level contour, and the contour shapes within the grid reflect local lexical tones, which are marked by the so-called "turning points". The grid could also expand and collapse, depending on whether there is a focus/non-focal position within a sentence. This is best summarized by Figure 2.4.

In recent years, there have also been quantitative, laboratory-based phonetics studies that attempt to model the layers of phonological components that affect the final Fo contour of an utterance. Xu, 2004, for example, proposed a model called "PENTA", which, among other things, described intonation and tones as parallel target approximation: each syllable carries two pitch targets, namely, one for tone and one for intonation/focus, which are realised simultaneously under the constraints of articulatory-timing. Another perception study done by Yuan and Shih, 2004 confirms the notion that sentence-level prosody

is realized through the raising/falling of the global pitch range and the raising/falling of the global pitch range of the terminal contour, while preserving the contour of individual lexical tones. Neurolinguistic research on the brain's processing of tones and intonation also shows that, when listeners are presented with lexical information, their brains show heightened sensitivity to a change in intonation than when no lexical information is available (Wang et al., 2023). Specifically, suppose a word is pronounced repeatedly with different intonational patterns. In that case, their brains show more neurological reaction to changes in intonational patterns compared to when they are presented only with a hummed version of the word.

To summarize, historical scholarship shows that sentence-level prosody in Mandarin, namely the surface Fo contour of a sentence, is determined by a layered juxtaposition of and interaction between lexical tones, global intonational patterns, and paralinguistic factors such as focus, emotion, and pragmatic intents. The effects of intonational patterns, in particular, consist of a global raising/lowering of the lexical tones with their contours intact, while simultaneously altering the contour of the terminal lexical tone. This indicates that prosody in Mandarin is layered and therefore needs to be examined as such.

2.2.4 Pragmatics and Prosody in Mandarin

Like many other languages, prosody in Mandarin is used for encoding pragmatic information and paralinguistic information such as emotion and attitude. Y. R. Chao, 1968 identified thirteen distinct kinds of pragmatic and paralinguistic functions that could be achieved through changes in prosodic patterns, though he does not claim to have an exhaustive list of all possible prosodic realizations of pragmatic and paralinguistic functions. As promising as it might sound, what Chao did in his work was merely an impressionistic description of the prosodic patterns of pragmatic and paralinguistic functions, which he was able to observe. Hence, there was no thorough phonetic description for what those prosodic patterns would entail phonetically and phonologically. The pragmatic functions that have the most thorough phonetic descriptions are questions, statements, and commands.

Given the fact that the very nature of prosody in Mandarin is one layered with many different levels of components, it is challenging to give a single, definitive description of what the Fo contours of statements in Mandarin look like. It is even harder to unpack the pragmatic details of a statement, and some question whether it is accurate to propose such a pragmatic category at all. For simplicity's sake, here a statement shall be defined as an attitude-neutral, informative sentence. Generally speaking, statements in Mandarin

generally consist of Fo contours that have not been altered by sentence-level intonation (White, 1980). In other words, the kind of global pitch-level raising that would happen under frameworks such as grid Gårding, 1987, 1993 and PENTA (Xu, 2004) would not occur. Aside from the sentence not having a global pitch-level raise, there is also a noticeable drop at terminal syllables (Y. R. Chao, 1968; Li & Thompson, 1989; White, 1980).

The picture becomes even less clear, however, when it comes to questions in Mandarin. Before a discussion could be made about the prosody of questions, a brief overview of the pragmatics of questions in Mandarin is needed.

Historically, questions in Mandarin have been described based on their syntactic constructions and pragmatic functions. Li and Thompson, 1989, for example, proposed a four category paradigm for Mandarin questions: question-word-question, disjunctive question, tag question, and particle question. Lee, 2005, on the other hand, gave an account which included questions that end with a grammatical particle, a category which he referred to as "grammatically marked questions", and questions that do not end with a grammatical particle and whose pragmatic functions are solely encoded in the prosodic domain, a category which he referred to as "grammatically unmarked questions" [6] . Under grammatically unmarked questions, there are two subcategories of questions mentioned by Lee, 2005, which he referred to as "Echo-Questions" and "InfoSeek-Questions". According to Lee, 2005, an echo-question is an interrogative sentence that is often the result of incredulity and surprise. In other words, it is used by one of the interlocutors in a conversation to seek confirmation for something that is hard to believe or out of the ordinary. Phonologically, these grammatically unmarked questions are formed directly on the basis of a statement, with the only difference being their intonational contour. Lee, 2005 further divided this subtype of questions into two pragmatic subtypes. One is what he calls the "info-seeking" kind where the locutor questions whether a statement is factually true. The other kind is what he calls "Echo-question", which serves to confirm whether a previously made statement is indeed factually correct. The difference in pragmatics between these two subtypes of questions is subtle, albeit noticeably different. Examples of these two types of questions look like the following:

1. 他今天回家

[6] See 2.1.2 where this dichotomy was briefly discussed

$t^ha˥$ $tɕin˥$ $t^hiən˥$ $hui˧˥$ $tɕə˧$
He now day go home

Is he going home today?

1. 他今天回家

$t^ha˥$ $tɕin˥$ $t^hiən˥$ $hui˧˥$ $tɕə˧˥$
He now day go home

Is he (really) going home today? (surprised)

A cursory look at the phonological transcription of these two sentences might not reveal many differences. The thing that one should notice, however, is that the terminal contour of the last morpheme/syllable of the second sentence instance is described to be noticeably higher than that of the first sentence instance. One of these sentences, therefore, is used by the speaker in question to indicate that he/she is trying to receive confirmation for a proposition that is included in the inquiry. The other one, however, while still technically an inquiry for confirmation, is one that was made with feelings of incredulity and surprise. Lee, 2005 argues that there are noticeable differences between their intonational patterns. First, the fact that both of them are not grammatically-marked (without any grammatical particle) means that their discourse function is realized intonationally, which Lee, 2005 had found to be encoded in an expanded pitch contour at the terminal position of the sentence contour. In addition, Lee, 2005 also noted that such expansion of pitch contour at the terminal position is more salient in echo-questions, compared to Infoseek-questions. It is unclear, however, whether such pragmatic information is encoded locally (namely within a segment of the contour) or globally (across the entire contour), as Lee, 2005 had found evidence for both types of explanations, even for grammatically unmarked echo-questions and infoseek-questions alone.

What is clear, based on Lee's research, however, is that native speakers of Mandarin utilize both a global pitch raise and a localized Fo cue for signaling questions. The issue then is when and how global pitch raise is used, and for what types of questions it is used.

Aside from questions and statements, another type of intonational function that is often utilized by speakers is an imperative, or what some might refer to as a command. Historically, there has not been much research done on the topic of prosody and intonation with regards to imperatives and commands in Mandarin. Y. R. Chao, 1968 did imply in his primer, however, that an imperative in Mandarin often implies a falling intonation at the terminal position of the contour. White, 1980 also briefly gave an

impressionistic description of what commands/imperatives should sound like in Mandarin, where she claims that a command in Mandarin usually involves a globally raised pitch level, plus a sharp drop/falling contour at the terminal position. There are, however, rebuttals to both Chao's and White's claims, as a recent study on rudeness and politeness in Mandarin and their relationship with prosody indicates that perceived rudeness in Mandarin has more to do with the speed of an utterance than with the Fo contour of a sentence (Fan & Gu, 2016).

2.3 The Acquisition of Intonation and Prosody by L2 Speakers

The relationship between prosody and pragmatics in English is much clearer and has been described in much more detail compared to Mandarin. Pierrehumbert, 1980, historically, was the first scholar to propose the concept of a "nuclear tone", which included the H/L distinction (High vs Low). A nuclear rise, therefore, is characterized as a question intonation in yes/no questions in English (Ladd, 2008; Pierrehumbert, 1980). In other words, a non-wh question would have a rising contour at the terminal position of the Fo contour of the sentence. A statement, on the other hand, would include a falling nuclear tone at the terminal position of the sentence contour. An echo question would involve a sharper, more acute rise at the terminal position, and a command would involve a steeper fall (Gussenhoven, 2004; Pierrehumbert, 1980). While there is no doubt that there are finer distinctions that could be made with regards to discussions about intonation and prosody in English. For simplicity's sake, I only focus on these broadly agreed-upon patterns with regards to intonation and prosody in English.

One thing that one might immediately notice, however, is that based on the above descriptions, English does not seem to have a globally realized Fo cue for realizing different intonations, and this is going to cause problems for the acquisition of intonation and prosody when learning another language. Historically, there has always been a paucity in the amount and scope of research and scholarship in general in this subfield of Second Language Acquisition. When it comes to L2 speakers' perception of intonation, the amount of research available is even sparser. There are, however, a handful of past research projects that are directly related to Second Language Acquisition but are nonetheless valuable. Liang and van Heuven, 2007, for example, demonstrated that speakers of Uyghur (a East-Turkic language spoken in Xinjiang, China) who have studied Mandarin as their L2 consistently outperformed Mandarin speakers

at identifying intonational cues at sentential endings, while they underperformed at identifying lexical tones when compared against native Mandarin speakers and speakers of other Sinitic languages [7]. The only explicit perceptual study on L2 intonation I was able to find was Yang and Chan, 2010. They did a perception experiment on American L2 learners of Mandarin, in which they asked them to identify questions and statements based on what they heard. In particular, Yang and Chan's study focused on American English speakers' perception of pragmatic cues that are encoded phonologically within the sentential endings of utterances and the interaction between lexical tones and pragmatic intent encoded using intonational contours. Yang and Chan found that while question intonations are, across the board, harder to distinguish than declaratives/statements, it gets even harder when the boundary tone interferes with the intonational target. Specifically, if someone were to say a sentence like the following:

1. 他今天放假

   $t^ha˥$ tɕin˥ $t^hiən˥$ faŋ˨˩ tɕə˨˩
   He now day have holiday

   He's going to be on holiday starting today.

This sentence, if formed into a grammatically unmarked question, as according to Lee, 2005, would be incredibly challenging for American L2 speakers to detect the question intonation, as the intonational target would be in direct conflict against the lexical tonal contour of the last character "假" (meaning "holiday"), which is pronounced with a falling tone as indicated by the IPA tone mark above. Another study done by Yuan, 2004, while being one done on native speakers, shows that question intonations are still harder to grasp and perceive, compared to that of statements.

In all, in this literature review, I have shown several things that would warrant a study such as this. First, Mandarin is poorly described in terms of its intonation and prosody, specifically the parts of its intonation and prosody that are used for pragmatic and emotional functions. Thus more research on Mandarin intonation and prosody is merited. Second, although there is still ongoing debate, Mandarin is widely believed to utilize two major strategies for achieving intonational prosody: an altered terminal position contour, plus a globally raised pitch level for the entire sentence's contour. Further research is needed to confirm and elaborate on these strategies. Third, English does not have such dual strategy for realizing its intonational prosody, as most such prosody is achieved by altering the contour of the terminal

[7] or so-called "dialects" in Sinitic Linguistics

position of a sentence. This suggests that L1 speakers of English may find it challenging to perceive the strategies used in Mandarin. Finally, this review has shown that very few studies have been done on L1 English L2 Mandarin speakers' perception of intonation and prosody, an issue of great interest to Mandarin language learners and teachers.

Chapter 3

Methodology and Materials

This study began with my observation that L1 American English speakers seem to struggle to understand and grasp question intonations in L2 Mandarin. While I did initially intend to measure directly L2 speakers' perception of intonational patterns by directly eliciting their responses for whether they think an utterance is a statement or question, I realized a similar research had already been done by Yang and Chan, 2010. I, therefore, decided to break down the intonational/pragmatic types that I had explored in the literature review (plain questions, emphatic questions, statements, and commands) into what I call sub-pragmatic components. I characterized each intonational/pragmatic type into two a priori binary sub-pragmatic components that I term "emphaticness" and "confidence." Emphaticness refers to the perceived emotional intensity of the speaker, while Confidence refers to the perceived assuredness of the speaker. According to this analytical scheme, a plain statement's pragmatic intent is judged to be composed of a perceived lack of emotional intensity (emphaticness) and a perceived sense of being assured (+confidence). An emphatic question, on the other hand, is judged to be composed of a perceived presence of both the speaker being unsure and emotionally intense. The following chart shows the paralinguistic encoding for each sentential variation:

Sentential Variations	Emphaticness	Confidence
Plain Statement	−	+
Plain Question	−	−
Emphatic Question	+	−
Command	+	+

Figure 3.1: Sub-pragmatic Coding for Each Sentential Variation

3.1 Research Question

I began this study wondering to what extent L1 English speaking L2 learners of Mandarin, a tonal language, can grasp sentential intonational contours which often function as subtle pragmatic cues. Specifically, I was interested in the extent to which they can perceive differences in a) emphaticness and b) confidence sub-pragmatic components of Mandarin utterances. To investigate this issue, I decided to compare these learners against two groups, "naïve" L1 English speakers with no previous exposure to tonal languages, and L1 speakers of Mandarin. This study seeks to answer the following two research questions: First, does L2 instruction in Mandarin make L1 speakers of English any more native-like in their abilities to detect emphaticness? Second, does L2 instruction in Mandarin make L1 speakers of English any more native-like in their perception of confidence in sentence intonation contours? Third, how close to native speakers are naïve speakers and L2 learners in judging the sub-pragmatic intent of Mandarin utterances from the sentence-level contours?

3.1.1 Stimuli

Stimuli used in this study were audio clips of sentences. These clips were four pragmatic variations of a single prototype sentence. The sentence clips that I used in this study were four pragmatic variations of a prototype sentence. These sentence variations were morphosyntactically identical to each other, as the only difference between them is their pragmatic function and, by extension, their intonational and prosodic profiles as well. These variations included plain statements, plain questions, emphatic questions, and commands. Plain questions and emphatic questions were based on Lee, 2005's typology of so-called

grammatically unmarked questions. A plain question is another way to phrase what Lee describes as an Infoseek-question, namely a question that is designed to seek information. An emphatic question was my equivalent of what Lee describes as an echo-question, namely, a question that is raised out of incredulity or surprise. A command is an imperative sentence that dictates an action. Contrary to plain and emphatic questions, I was not able to find detailed phonetic descriptions of what a command looks like in Mandarin, although Chao did give an impressionistic description, which states that a commanding voice should be realized with a falling intonation at the terminal contour of a sentence (Y. R. Chao, 1968), and it shall be described as such. Figures 3.2, 3.3, 3.4, 3.5 are graphs that illustrate the contour shapes of sentence-level prosodic patterns for all four types of sentence variations that are used as stimuli for this study

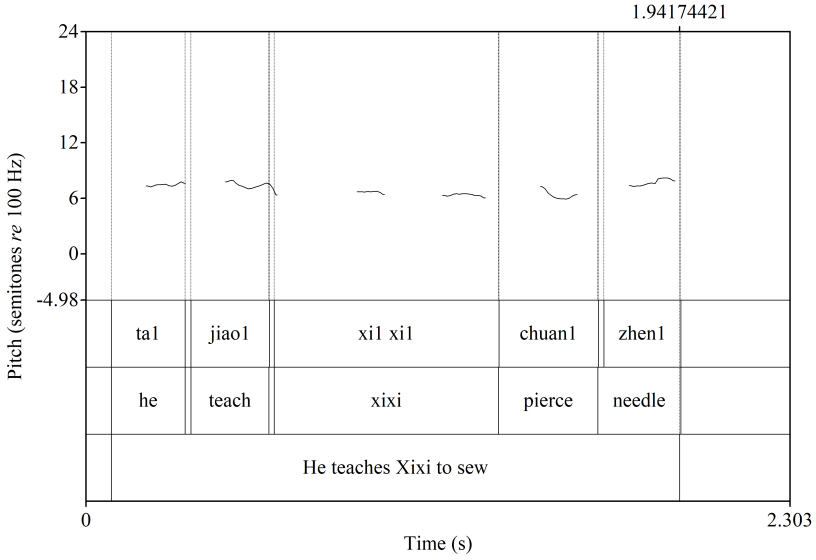

Figure 3.2: Plain Statement

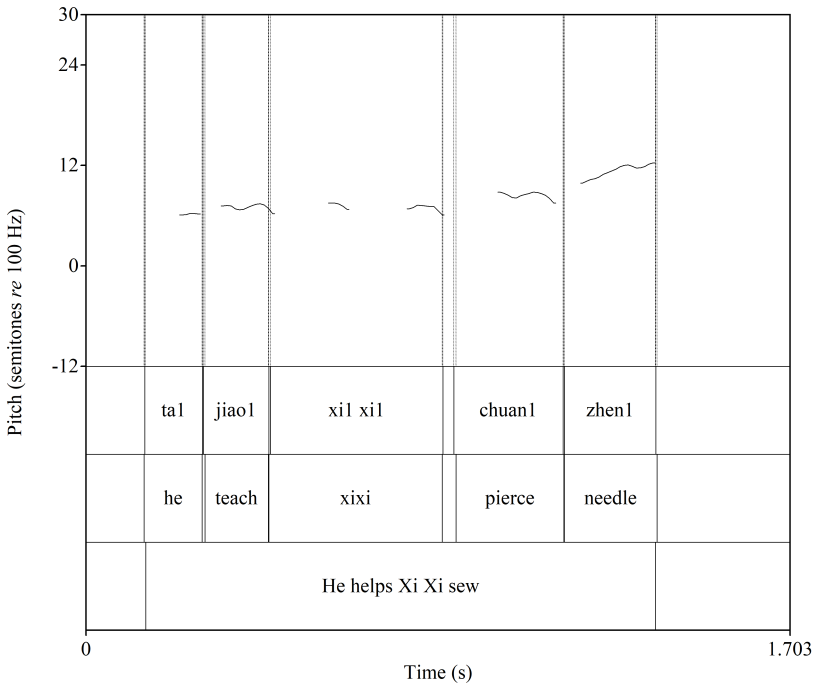

Figure 3.3: Plain Question (i.e., Infoseek-Question, according to Lee, 2005)

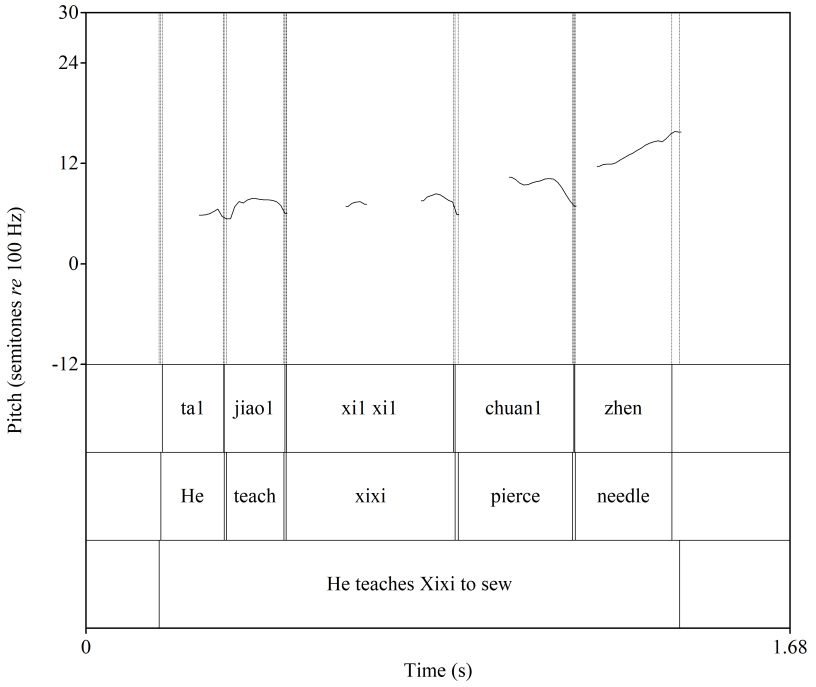

Figure 3.4: Emphatic Question (i.e., Echo-Question, according to Lee, 2005)

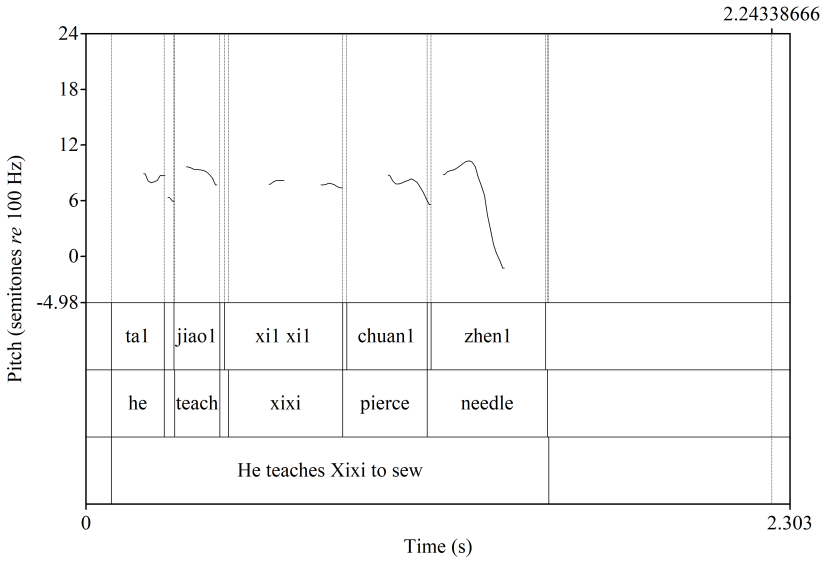

Figure 3.5: Command

3.2 Study Design

To answer the research questions, I designed an online perceptual study using qualtrics. First, I created ten sentence prototypes in Mandarin. For each sentence prototype, I created and recorded four pragmatic variations. These pragmatic variations included a plain statement, a plain question, an emphatic question, and a command. A pragmatic variation, in this context, is an audio clip that includes the exact same morphosyntactic combination of characters as the others, with the only difference being its intonational contour (See Appendix B for the IPA representation for each variation). For each sentential variation, I coded two paralinguistic, sub-pragmatic variables, which include "emphaticness" and "confidence". The coding of the variables was done a priori, based on the judgments of native speakers.

In order to control for unpredictable effects caused by tonal sandhi[1], I chose to only include sentences whose terminal contours have a level tone[2]. This level tone, depending on which discourse marker with which it interacts, either gets a flat, a rising, an acute rising, or a falling contour. To control for the effects of gender, I chose to include both a male and a female native speaker to record for each sentential variation. For each sentence variation that a participant was asked to rate, a female/male version was randomly chosen.

3.3 Participants

I recruited participants to represent people who are native speakers of a tonal language (native Mandarin speakers), those who have no experience with tonal languages whatsoever (naive English speakers), and those who have some experience with tonal languages (L2). In total, I was able to recruit 37 participants from three groups for the study. There were 8 English L1 Mandarin L2 speakers, 8 Mandarin L1 speakers, and 21 naive English L1 speakers. The participants of this study were either students who were enrolled at the University of Georgia, or people who lived in adjacent communities in the vicinity of Athens, Georgia.

[1] change of the phonetic realization of a tone based on its phonological environments. A brief introduction to tone sandhi could be found in Y. R. Chao, 1948's primer. I do not intend to discuss this at length in the proper text as this is irrelevant to the study

[2] as in ma˥, which is often recognized as the first tone in standard Mandarin tonology

3.4 Study Procedure

The recordings were distributed across a qualtrics survey that participants accessed by either clicking on a weblink or scanning a QR code. Participants were first asked to take a linguistic background survey in which they answered questions about their proficiency in Mandarin. The following shows the linguistic questionnaires that they have answered to:

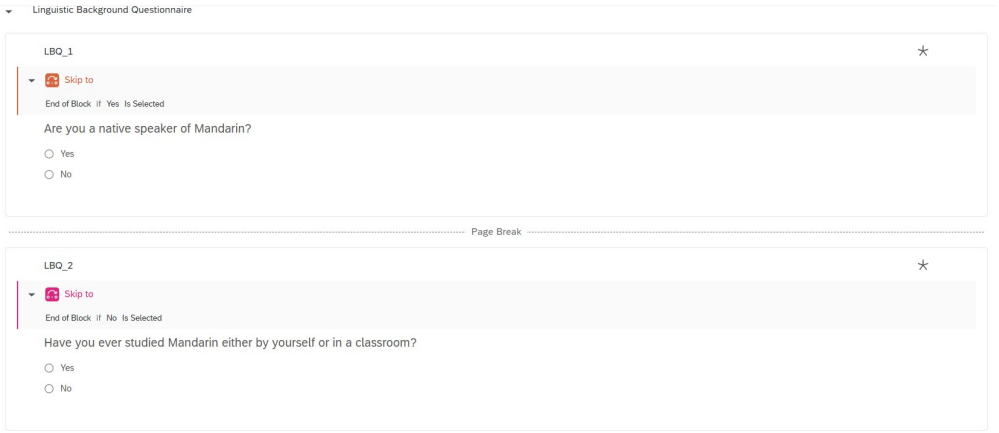

Figure 3.6: Linguistic Background Questionnaire 1

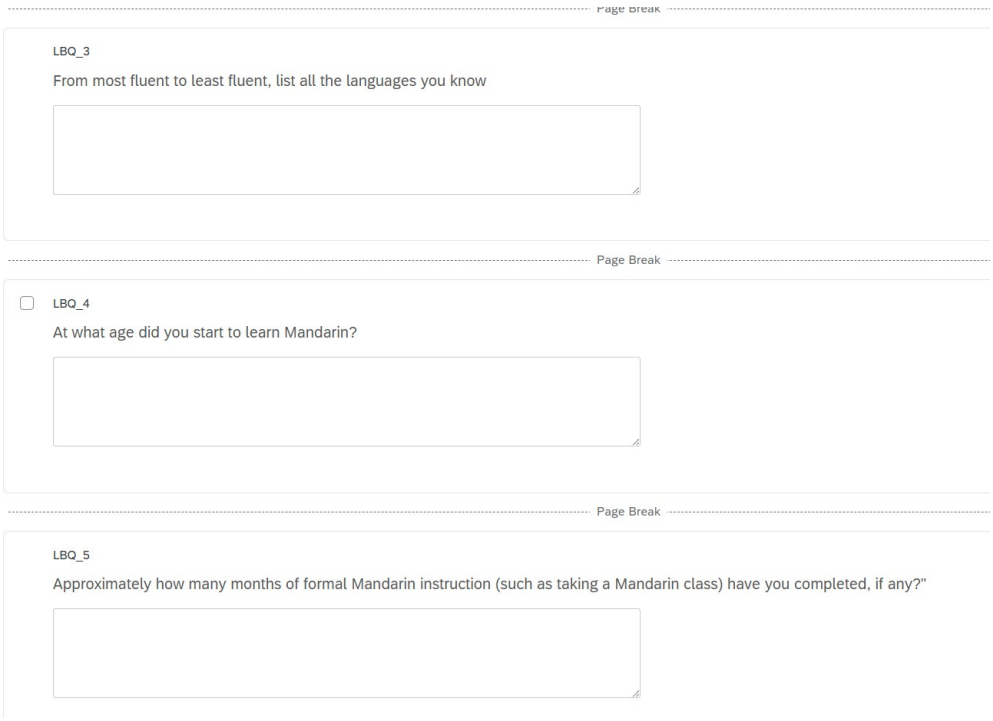

Figure 3.7: Linguistic Background Questionnaire 2

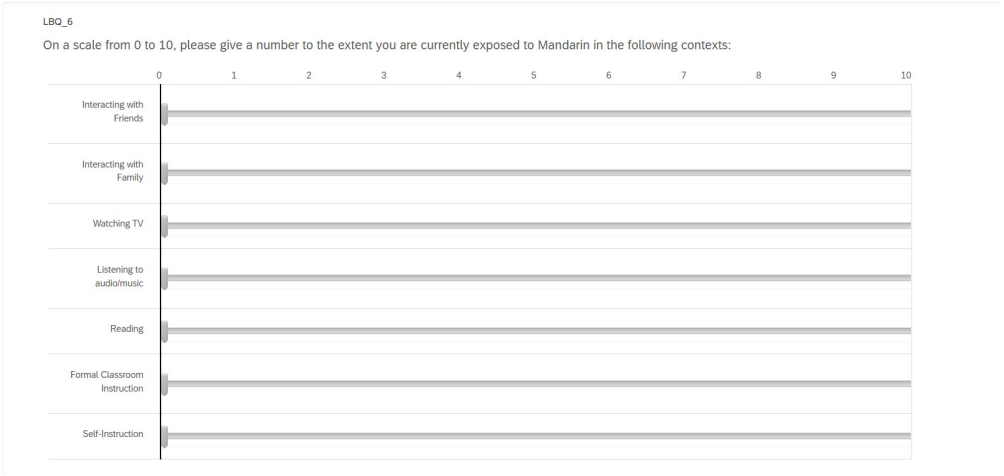

Figure 3.8: Linguistic Background Questionnaire 3

They would then be directed to a randomized sequence and answered questions about what they thought the speaker sounded like. They were presented with questions that elicited their responses for emphaticness and confidence ratings:

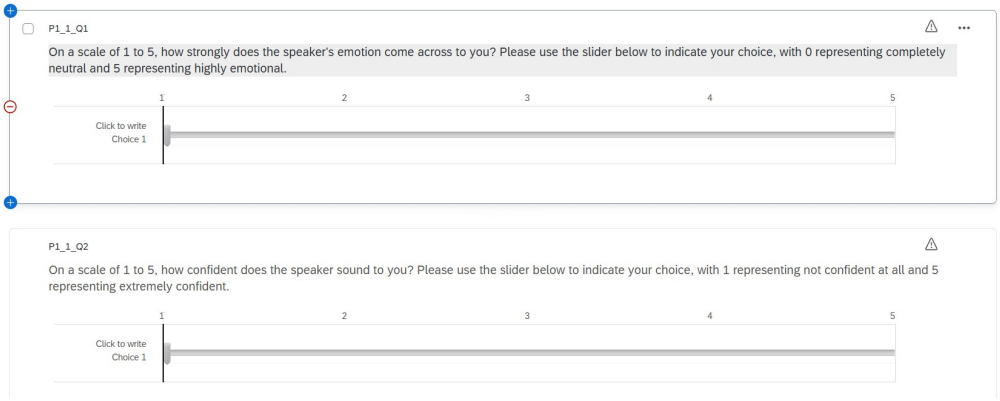

Figure 3.9: Emphaticness and Confidence Rating Questions

and questions that were irrelevant to the variables that I was interested in:

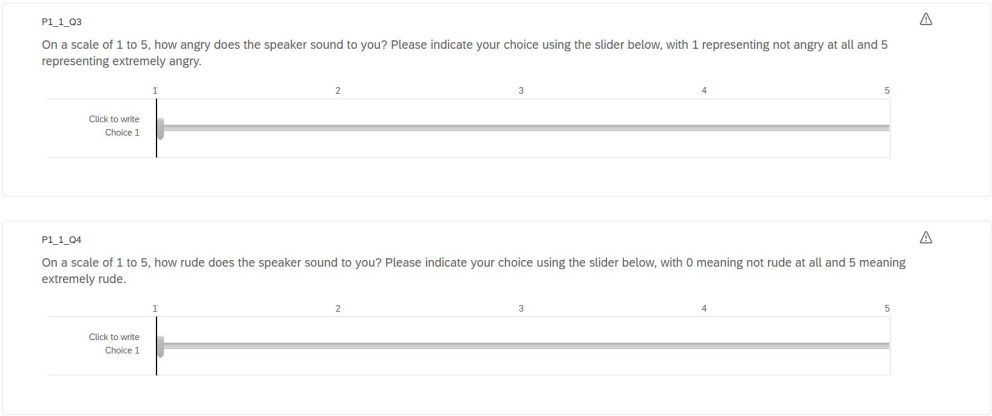

Figure 3.10: Filler Questions

Each participant gave emphaticness and confidence ratings for each sentential variation, which yielded a total of 1480 tokens per sub-pragmatic cue per participant for each sentence variation. After cleaning, there were 1470 usable rating tokens for each sub-pragmatic cue. All participants participated willingly and were informed of their rights to quit at any given time during the study.

To investigate the correlations between different speaker groups and their relative perceptual patterns of the encoded paralinguistic cues, I ran a linear mixed effects regression model for both the emphaticness rating and the confidence rating for each sentential variation token. Results and discussions will be presented in the next chapter (See chapter 4).

Chapter 4

Findings and Discussion

4.1 Mandarin Proficiency Analysis

The L2 participants of this study were primarily made up of UGA students who were taking Mandarin classes, plus one or two individuals who were not associated with the institution. For the UGA students, they were all from intermediate or upper-intermediate level classes. A linguistic background question was distributed to each participant during the study whose results were used later to calculate a proficiency score. At first, I intended to use this as a continuous variable in the quantitative modeling process. However, because of the paucity of the number of participants, the calculated proficiency score of the L2 participants appeared to be highly skewed towards highly proficient and those lacking in proficiency, making it much harder for this variable to become useful in the quantitative modeling process. As a result, proficiency score was not used as a variable in the quantitative modeling process, and L2 speakers were treated instead as a whole group.

4.2 Data Analysis

To analyze the data, I chose to fit a linear mixed-effects regression model for the rating score that the participants gave for each item. Participants and sentence prototypes are modeled as random effects. The narrator of the sentence clips is modeled as a fixed effect. The reason why I chose to do so is because historically, it has been well established that women, in general, are more likely to be perceived as less

authoritative and confident in perceptual studies like this. Since I hypothesized that there would be differences in how the three speaker groups would rate confidence and emphaticness for each sentential variation, they were coded as having an interaction in the model. In addition, since I also hypothesized that L2 speakers would both perform similarly and differently compared to naive and native speakers, I choose to designate them as the baselin in the modeling. In other words, their results were being compared to as the baseline. A summary of the two best models for the confidence rating and emphaticness rating is listed in Appendix A (See regression table A.1 and table A.2).

4.2.1 Emphaticness

Commands, emphatic questions, and plain questions showed significant differences in ratings of emphaticness when compared to plain statements. The coefficients for commands and emphatic questions matched with their assigned a priori emphaticness coding. The emphaticness ratings for plain questions, however, unexpectedly turned out to be higher than plain statements ($\beta = 0.942, p < .001$). For the listener type x sentential variation interaction items, the reference levels for comparison were coded as plain statement and L2, respectively, for sentential variation and speaker group. Therefore, each coefficient would reflect the additional effect of a given sentential type and speaker group combination relative to L2 speakers' rating of plain statement type sentences.

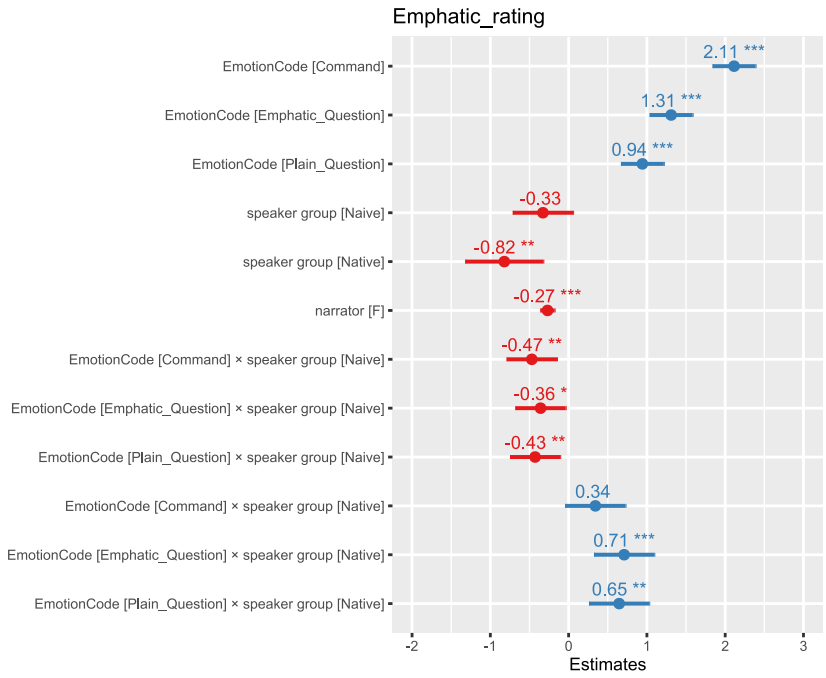

Figure 4.1: Model Summary for Emphaticness Ratings

Results showed that for naïve speakers, utterances coded as command, emphatic question, and plain question all had significantly negative effects relative to L2 speakers' rating of plain statements ($\beta = -0.470, p < .001$; $\beta = -0.358, p < .01$; $\beta = -0.428, p < .001$, respectively).

Specifically, command sentences ($\beta = -0.470, p < .01$), emphatic questions ($\beta = -0.358, p < .01$), and plain questions ($\beta = -0.428, p < .001$) were all rated as significantly less emphatic by naïve speakers.

Conversely, native speakers showed an opposite pattern: for commands($\beta = 0.342, p < .05$), emphatic questions($\beta = 0.709, p < .001$), and plain questions ($\beta = 0.646, p < .001$), they were all rated as more emphatic by native speakers than they were rated by L2 speakers.

Overall, for command and emphatic questions, their ratings were shown to have a strong interaction with regards to speaker group. The coefficients of the interactions correlated with the a priori assigned emotional coding of the sentential variations. For plain questions, however, it seems as though native speakers are likely to rate higher, which does not correspond to its assigned emotion code.

4.2.2 Confidence

Similar to emphaticness, speakers' rating of each stimuli item's perceived confidence score was also modeled after the aforementioned categorical variables, namely narrator, sentential type, speaker group, sentence prototype, and participant. For sentential variations, the model showed significant effects for command sentences ($\beta = 0.982, p < .001$), meaning that command sentences were rated higher across the board by all speaker groups when compared to plain statements. For the interaction part, there are only two things that are worth noting, namely native speakers' rating of emphatic questions and plain questions, when compared with the baseline (L2 speakers' rating of plain statements). Both interactions showed a significant negative effect ($\beta = -1.175, p < .001$, $\beta = -1.344, p < .001$, respectively). This goes against the a priori assigned emotional code for plain question and emphatic question, since both would have been expected to receive higher ratings compared to plain statements.

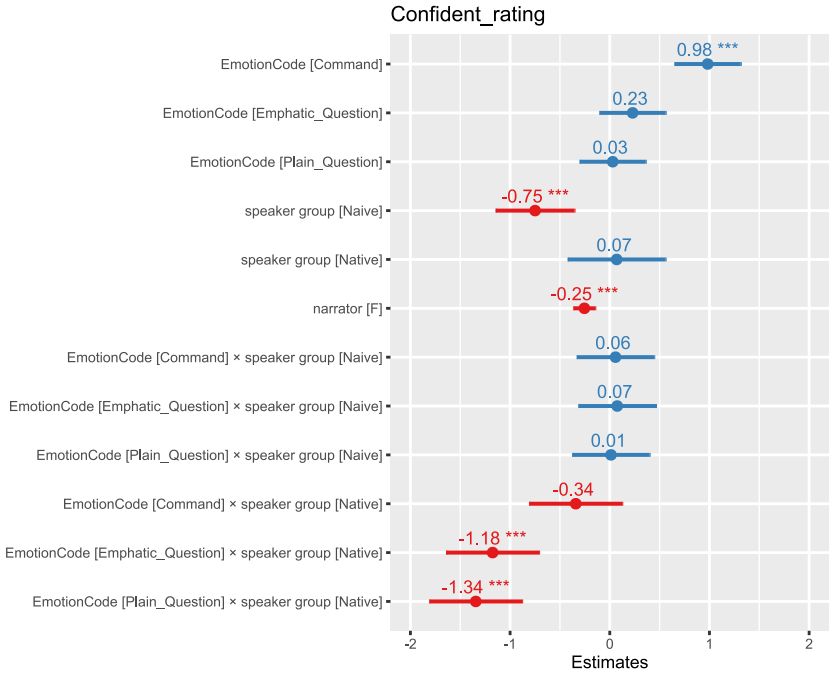

Figure 4.2: Model Summary for Confidence Rating

4.2.3 Emphaticness Results Discussion

Looking at our data, there are several things that are important and worth noting in our discussion. First, one could notice that sentential type (emotion code in the model) has a statistically significant effect on participants' rating of emphaticness for all sentence variations across the board. Commands, emphatic questions, and plain questions all appear to be rated higher with regards to emphaticness compared to plain statements. An unexpected result is that plain questions also appear to be rated higher for emphaticness compared to plain statements.

For commands and emphatic questions, there was a statistically significant interaction between sentence variation and speaker group. Specifically, naive monolingual English speakers appeared to rate emphaticness lower compared to L2 speakers, and native Mandarin speakers appeared to rate emphaticness higher for them compared to L2 speakers. This is an interesting and important finding, as commands and emphatic questions are inherently more emphatic variations compared to plain statements. The fact that the L2 speakers in this study appeared to rate emphaticness higher compared to naive monolingual English speakers for these two sentence variations but relatively lower compared to native Mandarin speakers suggest that L2 speakers appear to have acquired, at the very least, partial knowledge of what it means to have an intonational/prosodic contour that signifies a fortition of one's emotion and pragmatic intent. In other words, L2 speakers' emphaticness ratings for emphatic sentence variations appear to be statistically different from those of naive speakers, but not so much that there is no difference between them and the natives. There is still a noticeable gap (the fact that there is a statistically significant difference between L2 ratings vs native ratings) that separates L2 speakers from native speakers with regards to their ratings of emphaticness when they are presented with an emphatic sentence.

As was mentioned earlier, it is interesting that plain questions also appear to be rated higher with regards to emphaticness, compared to plain statements. This is interesting because it makes one wonder if emphaticness is properly being interpreted by the participants. It is entirely possible that some of the participants interpreted it as encompassing all kinds of emotions, such as hesitation or showing a desire for inquiry. If such were the case, then it is questionable whether the interaction effect for emphatic question and speaker group was indeed caused by L2 speakers' more accurate perception of pragmatic prosody. It is entirely possible that they could have interpreted hesitation as emphaticness, which is supposed to be a separate pragmatic cue that denotes a question rather than emphatically made utterance. The coefficients

for the interaction effects, however, tell us that even if such were the case, emotional/pragmatic prosody with regards to emphaticness was picked up by L2 speakers, as the coefficients for emphatic questions were still higher compared to the coefficients for plain questions.

Aside from these statistically significant effects, a more straightforward way to visualize our model is through an emmeans graph. An emmeans graph is generated through calculating the emmeans of ratings with regards to speaker group and sentential variations. Essentially, it makes predictions based on the existing coefficients and parameters of a linear model. The following is an emmeans graph for the emphaticness rating model:

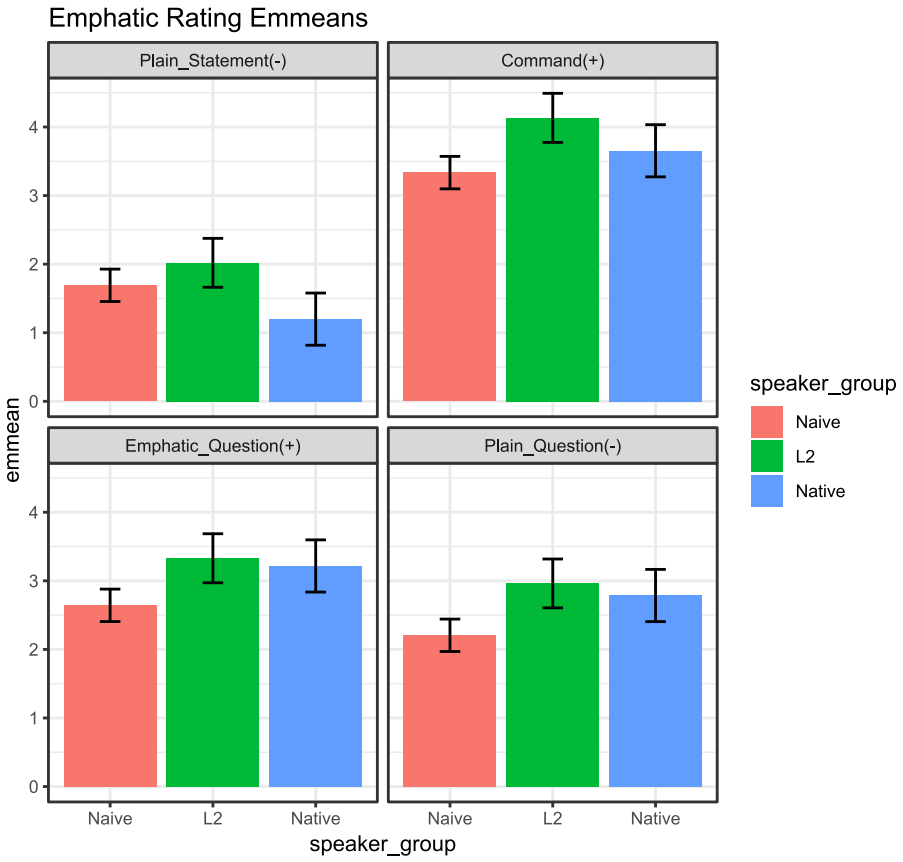

Figure 4.3: Emphatic Rating by Speakers

One thing that one could immediately notice is that for some reason, both L2 and naive monolingual English speakers' ratings of plain statements look much higher than those made by native speakers. I am not sure why this is happening. One explanation that is semi-plausible, however, is based on the

assumption that Mandarin Chinese speakers have a wider pitch range when compared to English speakers (White, 1980). The fact that Mandarin speakers have a higher pitch-range when compared to English speakers could have several implications. Because Mandarin speakers are required to achieve both lexical tone and intonational targets (Y. R. Chao, 1968; Gårding, 1979, 1987; Xu, 2004), it could be that their more varied and relatively more sudden changes in pitch/Fo is interpreted as emphaticness rather than what they actually are.

4.2.4 Confidence Results Discussion

Several conclusions could be drawn, based on our quantitative modeling and the parameters of the model. First, one could notice that Command sentences were rated, across the board (namely, by all members of our three speaker groups), as more confident-sounding than plain statements. This is, of course, to be expected, as commands/imperatives are cross-linguistically perceived to be more authoritative-sounding and, by extension, more confident than a plain statement that is only intended to inform Searle, 1979.

Second, it is interesting to note that there is a significant effect of the speaker group when it comes to participants' rating of confidence across the board. As one can see, the naive, monolingual English speakers appear to rate all sentences as slightly less confident when compared to L2 speakers. It is not clear what may have caused this. One speculation that one could make about this is taken from White, 1980, who suggested that Mandarin has a broader pitch range compared to English when it comes to the prosody and intonation of these two languages, and who also hinted at the notion that English speakers tend to perceive more varied pitch ranges as hesitation. If that is so, then this finding in our model would serve as evidence affirming not only that English speakers perceive prosody fundamentally differently, but also that each language utilizes prosody and intonation differently in terms of emotion and pragmatics.

What is more important and worth noting in our finding, however, is that there is a significant effect of interaction between speaker group and emotion code with regards to how emphatic questions and plain questions were rated by native Mandarin speakers. When asked to rate the perceived confidence of the speaker of a sentence clip, native Mandarin speakers appear to rate both emphatic questions and plain questions as less confident, compared to both plain statements and how L2 speakers rate emphatic questions and plain questions. In other words, for sentence variations that should be perceived as less confident, native Mandarin speakers rated them lower than L2 speakers. This finding is significant because

it shows that, for whatever reason, the intonational and prosodic contours of emphatic and plain questions were not perceived to be as denoting hesitation as how they were perceived by native speakers. On the other hand, since we do not have a universally agreed-upon framework for describing the prosody of questions in Mandarin, we can only assume that a certain part of the prosodic profile of our emphatic and plain questions made the identification of this particular pragmatic cue (namely, confidence) more challenging for L2 speakers. Presumably, since all existing frameworks assume that there is terminal rising for questions in Mandarin, regardless if they assume global pitch-raising, we can conclude that, at the very least, this terminal rising is either not being properly interpreted by L2 speakers as a pragmatic cue for when there is hesitation, or that it is being interfered by other prosodic components of the sentence that messes with their perception. Figure 4.4 is an emmeans graph based on our model.

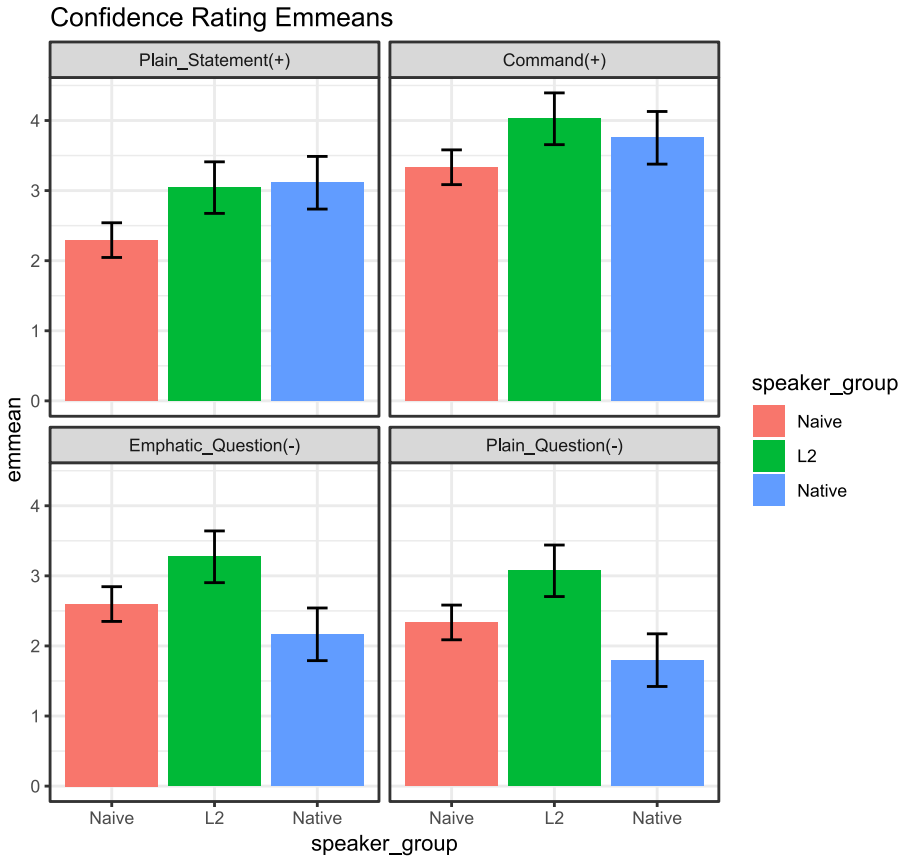

Figure 4.4: Confident Rating by Speakers

Aside from the effects that have been shown to be statistically significant in our model, the emmeans graph for our data suggests interesting patterns that are not directly verifiable in the model alone. First, one could notice that, although not statistically significant, L2 speakers' ratings of confidence appear to be closer to those of native speakers for plain statements and commands. This means that, for sentences whose terminal contours are either neutral or falling, L2 speakers appear to be able to register such prosodic correlates as a pragmatic/emotional cue for confidence as well as native speakers of Mandarin do. It is important, however, that this claim not be blown out of proportion, as this is only based on a superficial examination of a graph which itself is based on predicated values of a linear model, and we have already established that these observed patterns in our emmeans graph do not appear to be statistically significant in the model.

4.3 Summary

Overall, the model shows significant interaction effects between speaker group and sentence variation types for both emphaticness and confidence ratings. For emphaticness ratings, L2 speakers appear to have acquired some prosodic/intonational knowledge of what emphaticness means in Mandarin. This acquisition, however, appears to be partial, as their ratings appear to be still statistically lower than those made by native Mandarin speakers for sentence variations that should be deemed as emphatic. For confidence ratings, L2 speakers would still seem to struggle with the acquisition of an understanding of intonational and prosodic patterns used for signaling inquery, as native speakers showed statistically lower ratings with regards to confidence for both types of questions than L2 speakers did.

Chapter 5

Conclusion

While how prosody works in Mandarin exactly remains a question to be answered, this study was able to confirm that L2 Mandarin speakers do behave differently when they are asked to rate their perception of pragmatic and emotional cues. Specifically, the results from the linear model for confidence rating in this study have helped solidify the notion that question intonation in Mandarin is indeed challenging for learners to acquire. The results from the linear model for emphaticness rating, on the other hand, showed that L2 speakers were able to acquire, at the very least, some knowledge of the intonational/prosodic patterns that signify emphaticness in a sentence. This conclusion is drawn from the fact that the model shows that naïve, monolingual English speakers consistently rated sentences with the emphaticness coding—namely, sentences that are supposed to sound emphatic and filled with emotion—lower compared to L2 speakers. However, it is also true that native speakers' ratings of emphaticness for emphatic questions are statistically higher compared to those of L2 speakers. This could suggest that, at least based on the results of the model alone, L2 speakers were able to perceive emphaticness in a way closer to how native speakers perceive it, though it remains unclear as to what cognitive/linguistic strategies they employed to reach such ratings. This is the result of one of the drawbacks of this study, which shall be discussed in Section 5.1.

Since the debate between local vs global intonational cue theories in Mandarin prosody is still not resolved, there is no cogent framework with which one can reasonably work out a theory as to how exactly the Fo contours of our sample sentences were perceived. Specifically, did our participants make their judgments based on their perception of the terminal contour? Did they make their judgments based on a more global sentence-level intonational pattern? These are questions that remain to be answered, which

cannot be properly addressed by our current study. And the fact that we cannot be sure whether they used it and, if they did, what intonational cues they used at all means that any claim we could make based on our model should be analyzed as is.

5.1 Drawbacks and Limitations

One of the drawbacks of this study is the lack of a universal consensus on how prosody works in Mandarin. The fact that we have scholars from both sides (namely, the local pitch change vs global pitch change camps when it comes to pragmatic intonations in Mandarin) presenting evidence that is equally convincing means that we are presented with a stalemate concerning our theoretical basis for the study. There is simply not one workable theory with which we can properly account for our observations and make conclusions regarding how L2 learners localize and identify both intonational and prosodic cues. Because of this, this study was only able to examine L2 speakers' perception of Mandarin intonation and their understanding of intentionally realized pragmatic functions on a macro-level. In other words, this study does not tell us—at the very least, not conclusively—whether learners behaved more/less like native speakers in their ratings because they correctly identified the exact phonological correlates of the emotional/prosodic cues in question. It does not negate the possibility that their decision was affected by factors that are 1. external to said phonological correlates or 2. not related to the phonology of prosody at all. That being said, the fact that our sentences were all morphosyntactically identical suggests that it is still very likely that our participants made their judgments based on some kind of intonational, prosodic correlates, though it remains unclear what said intonational and prosodic correlates were.

The other drawback that deserves some attention is my intentional decision not to use proficiency scores as a variable in the quantitative modeling. This was mainly caused by the number of people who were in the pool of participants. For L2 speakers, I was only able to recruit 8. The majority of those speakers were students from different levels of Chinese classes at the University of Georgia, with a few speakers who were unaffiliated with the institution. As such, there were simply not enough people for an even distribution of language proficiency in our pool of participants. Thus, we cannot conclude whether this perceived increase in awareness of some intonational/prosodic cues among our L2 speakers (compared to naïve speakers) is correlated with proficiency. The only conclusion we can draw is that L2

speakers, as a whole, showed more native-like perception for emphaticness and more naïve-like perception for confidence.

Thirdly, this study is also limited in that the format it takes does not guarantee that our participants were being exposed to the stimuli under similar conditions when they participated. While it is true that the stimuli were all presented in the form of audio clips being played in an online survey, there was no way to ensure that the participants were being attentive to the task. This means that while there are observed patterns in the results of this study, their validity should be compared to the results of a more controlled study and judged accordingly.

5.2 Summary and Direction for Future Studies

In all, this study demonstrates that L2 Mandarin speakers show some signs of being able to detect emotional and pragmatic cues in the language based on sentence-level prosody and intonation. L2 speakers in this study, as a whole, showed more native-like perception of emphaticness compared to naïve monolingual English speakers, while they showed less native-like perception of confidence when compared to native speakers. There needs to be further phonetic study on prosody and pragmatics in both Mandarin and English to make cross-linguistic comparisons more feasible and their results more meaningful.

Appendix A

Tables and Figures

Table A.1: Results: Emphaticness Rating

	Dependent variable:
	Emphatic_rating
EmotionCodeCommand	2.114^{***} (0.140)
EmotionCodeEmphatic_Question	1.310^{***} (0.140)
EmotionCodePlain_Question	0.942^{***} (0.140)
speaker_groupNaive	-0.328^{*} (0.196)
speaker_groupNative	-0.822^{***} (0.254)
narratorF	-0.270^{***} (0.046)
EmotionCodeCommand:speaker_groupNaive	-0.470^{***} (0.164)
EmotionCodeEmphatic_Question:speaker_groupNaive	-0.358^{**} (0.164)
EmotionCodePlain_Question:speaker_groupNaive	-0.428^{***} (0.163)
EmotionCodeCommand:speaker_groupNative	0.342^{*} (0.196)
EmotionCodeEmphatic_Question:speaker_groupNative	0.709^{***} (0.196)
EmotionCodePlain_Question:speaker_groupNative	0.646^{***} (0.196)
Constant	2.154^{***} (0.179)
Observations	1,470
Log Likelihood	−1,945.902
Akaike Inf. Crit.	3,923.803
Bayesian Inf. Crit.	4,008.491

*p<0.05; **p<0.01; ***p<0.001

Table A.2: Results: Confidence Rating

	Dependent variable:
	Confident_rating
EmotionCodeCommand	0.982*** (0.169)
EmotionCodeEmphatic_Question	0.229 (0.169)
EmotionCodePlain_Question	0.030 (0.169)
speaker_groupNaive	−0.749*** (0.202)
speaker_groupNative	0.069 (0.250)
narratorF	−0.255*** (0.056)
EmotionCodeCommand:speaker_groupNaive	0.058 (0.198)
EmotionCodeEmphatic_Question:speaker_groupNaive	0.075 (0.198)
EmotionCodePlain_Question:speaker_groupNaive	0.012 (0.198)
EmotionCodeCommand:speaker_groupNative	−0.341 (0.238)
EmotionCodeEmphatic_Question:speaker_groupNative	−1.175*** (0.238)
EmotionCodePlain_Question:speaker_groupNative	−1.344*** (0.237)
Constant	3.170*** (0.186)
Observations	1,470
Log Likelihood	−2,214.279
Akaike Inf. Crit.	4,460.558
Bayesian Inf. Crit.	4,545.246

*p<0.05; **p<0.01; ***p<0.001

Appendix B

Stimuli Items

(1) 孙英今天开飞机

a. Plain Statement

tɕin˥ tʰiən˥ sun˥ jiŋ˥ kʰai˥ fei˥ tɕi˥
now day Sunying operate fly.ADJ machine

Sunying is flying the airplane today.

b. Plain Question

tɕin˥ tʰiən˥ sun˥ jiŋ˥ kʰai˥ fei˥ tɕi˩
now day Sunying operate fly.ADJ machine

Is Sunying flying the plane today?

c. Emphatic Question

tɕin˥ tʰiən˥ sun˥ jiŋ˥ kʰai˥ fei˥ tɕi˩˧
now day Sunying operate fly.ADJ machine

(huh) Sunying is flying the plane today?

d. Command

tɕin˥ tʰiən˥ sun˥ jiŋ˥ kʰai˥ fei˥ tɕi˩˨
now day Sunying operate fly.ADJ machine

It's Sunying's turn to fly the airplane today!

(2) 他今天敲钟

a. Plain Statement

tʰa˥ tɕin˥ tʰiən˥ tɕʰao˥ dʒoŋ˥
He now day hit bell

He is ringing the bell today (It's his turn)

b. Plain Question

tʰa˥ tɕin˥ tʰiən˥ tɕʰao˥ dʒoŋ˧
He now day hit bell

He is ringing the bell today?

c. Emphatic Question

tʰa˥ tɕin˥ tʰiən˥ tɕʰao˥ dʒoŋ˧˩
He now day hit bell

He is ringing the bell today? (huh)

d. Command

tʰa˥ tɕin˥ tʰiən˥ tɕʰao˥ dʒoŋ˩
He now day hit bell

He is ringing the bell today! (It's his turn)

(3) 他教茜茜穿针

a. Plain Statement

tʰa ˥ tɕiao ˥ ɕi˥ɕi˥ tʃuən˥ dʒən˥
He teach XiXi pierce needle

He is teaching Xixi how to thread a needle

b. Plain Question

tʰa ˥ tɕiao ˥ ɕi˥ɕi˥ tʃuən˥ dʒən˧
He teach XiXi pierce needle

Is he teaching Xixi how to thread a needle?

c. Emphatic Question

tʰa ˥ tɕiao ˥ ɕi˥ɕi˥ tʃuən˥ dʒən˧˩
He teach XiXi pierce needle

He is teaching Xixi how to thread a needle? (Huh)

d. Command

tʰa ˥ tɕiao ˥ ɕi˥ɕi˥ tʃuən˥ dʒən˩
He teach XiXi pierce needle

He is teaching Xixi how to thread a needle (He's doing it!)

(4) 他发出轻轻的声音

a. Plain Statement

tʰa˥ fa˥ tʃu˥ tɕiŋ˥ tɕiŋ˥ tə˩ ʃəŋ˥ jin˥
he make soft-soft poss sound

He's making a very soft sound

b. Plain Question

tʰa˥ fa˥ tʃu˥ tɕiŋ˥ tɕiŋ˥ tə˩ ʃəŋ˥ jin˧
he make soft-soft poss sound

Is he making a very soft sound?

c. Emphatic Question

tʰa˥ fa˥ tʃu˥ tɕiŋ˥ tɕiŋ˥ tə˩ ʃəŋ˥ jin˧
he make soft-soft poss sound

Is he (really) making a very soft sound?

d. Command/Emphasis

tʰa˥ fa˥ tʃu˥ tɕiŋ˥ tɕiŋ˥ tə˩ ʃəŋ˥ jin˧˥
he make soft-soft poss sound

He's making a very soft sound!

(5) 他帮妈妈刷锅

a. Plain statement

tʰa ˥ paŋ ˥ ma ˥ ma ˥ ʃua˥ guo˥
he help (his)mom clean wok

He helps his mom with cleaning the wok.

b. Plain Question

tʰa ˥ paŋ ˥ ma ˥ ma ˥ ʃua˥ guo˧
he help (his)mom clean wok

Does he help his mom with cleaning the wok

c. Emphatic Question

tʰa ˥ paŋ ˥ ma ˥ ma ˥ ʃua˥ guo˧
he help (his)mom clean wok

He is the one helping his mom with cleaning the wok?

d. Command

tʰa ˥ paŋ ˥ ma ˥ ma ˥ ʃua˥ guo˥˘
he help (his)mom clean wok

He is helping his mom with cleaning the wok!

(6) 你回我家拿东西

a. Plain Statement

ni ˧˥ hui ˧ wo ˧˥ tɕiə ˥ na ˧ tong ˥ ɕi ˥
you go-back my house take something

You are coming back to my house to pick up something

b. Plain Question

ni ˧˥ hui ˧ wo ˧˥ tɕiə ˥ na ˧ tong ˥ ɕi ˧
you go-back my house take something

Are you coming back to my house to pick up something?

c. Emphatic Question

ni ˧˥ hui ˧ wo ˧˥ tɕiə ˥ na ˧ tong ˥ ɕi ˧
you go-back my house take something

Are you coming back to my house to pick up something? (huh)

d. Command

ni ˧˥ hui ˧ wo ˧˥ tɕiə ˥ na ˧ tong ˥ ɕi ˥˘
you go-back my house take something

Come back to my house and take something!

(7) 哥哥和姐姐过家家

a. Plain Statement

kə ˥ kə˥ hə ˧ tɕie ˧˥ tɕie ˧˥ kuo ˨˩ tɕiə ˥ tɕiə ˥
older-brother and older-sister play-house

My older brother and older sister are playing house together.

b. Plain Question

kə ˥ kə˥ hə ˧ tɕie ˧˥ tɕie ˧˥ kuo ˨˩ tɕiə ˥ tɕiə ˧
older-brother and older-sister play-house

Are my older brother and older sister playing house together?

c. Emphatic Question

kə ˧ kəˀ hə ˧ tɕie ˧ tɕie ˧ kuo ˨ tɕiə ˧ tɕiə ˧
older-brother and older-sister play-house

My older brother and older sister are playing house together?

d. Command

kə ˧ kəˀ hə ˧ tɕie ˧ tɕie ˧ kuo ˨ tɕiə ˧ tɕiə ˨˩
older-brother and older-sister play-house

My older brother and older sister are supposed to be playing house together!

(8) 他教茜茜唱情歌

a. Plain Statement

tʰa ˧ hə ˧ ɕi ˧ ɕi ˧ tʃaŋ ˨ tɕʰiŋ ˧ kə ˧
he and Xixi sing love-song
He and Xixi are singing love songs together

b. Plain Question

tʰa ˧ hə ˧ ɕi ˧ ɕi ˧ tʃaŋ ˨ tɕʰiŋ ˧ kə ˧
he and Xixi sing love-song
Is he and Xixi singing love songs together?

c. Emphatic Question

tʰa ˧ hə ˧ ɕi ˧ ɕi ˧ tʃaŋ ˨ tɕʰiŋ ˧ kə ˧
he and Xixi sing love-song
He and Xixi are singing love songs together? (Huh)

d. Command

tʰa ˧ hə ˧ ɕi ˧ ɕi ˧ tʃaŋ ˨ tɕʰiŋ ˧ kə ˨˩
he and Xixi sing love-song
He and Xixi are supposed to be singing love songs together!

(9) 他教茜茜学音标

a. Plain Statement

tʰa˧ tɕiao˧ ɕi˧ ɕi˧ ɕye˧ jin˧ piao˧
he teach Xixi study sound-symbol

He is teaching Xixi the phonetic alphabet.

b. Plain Question

tʰa˥ tɕiao˥ ɕi˥ ɕi˥ ɕye˧ jin˥ piao˧
he teach Xixi study sound-symbol

Is he teaching Xixi the phonetic alphabet?

c. Emphatic Question

tʰa˥ tɕiao˥ ɕi˥ ɕi˥ ɕye˧ jin˥ piao˧
he teach Xixi study sound-symbol

He is (really) teaching Xixi the phonetic alphabet?

d. Command

tʰa˥ tɕiao˥ ɕi˥ ɕi˥ ɕye˧ jin˥ piao˨
he teach Xixi study sound-symbol

He is teaching Xixi the phonetic alphabet!

(10) 哥哥和姐姐吃锅巴

a. Plain Statement

kə˥kə˥ hə˥ tɕie˥ tɕie˥ tʃɯ˥ ko˥ pa˥
brother(older) and sister(older) eat crispy-rice

(My) older brother and sister are eating crispy rice

b. Plain Question

kə˥kə˥ hə˥ tɕie˥ tɕie˥ tʃɯ˥ ko˥ pa˧
brother(older) and sister(older) eat crispy-rice

Are (My) older brother and sister eating crispy rice?

c. Emphatic Question

kə˥kə˥ hə˥ tɕie˥ tɕie˥ tʃɯ˥ ko˥ pa˧
brother(older) and sister(older) eat crispy-rice

Are (My) older brother and sister (really) eating crispy rice?

d. Command

kə˥kə˥ hə˥ tɕie˥ tɕie˥ tʃɯ˥ ko˥ pa˨
brother(older) and sister(older) eat crispy-rice

(My) older brother and sister are eating crispy rice

Literature Cited

Abercrombie, D. (1967). Elements of general phonetics edinburgh: Edinburgh university.

Austin, J. L. (1975). *How to do things with words*. Harvard university press.

Chang, N.-C. T. (1958). Tones and intonation in the chengtu dialect (szechuan, china). *Phonetica*, *2*(1-2), 59–85.

Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken chinese*. Harvard University Press.

Chao, Y. R. (1968). *A grammar of spoken chinese.* University of California Press.

Chao, Y.-R. (1933). Tone and intonation in chinese. *Bulletin of the Institute of History and Philology, Academia Sinica*, *4*, 121–134.

Chien, P.-J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2020). Neural correlates of intonation and lexical tone in tonal and non-tonal language speakers. *Human brain mapping*, *41*(7), 1842–1858.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, *40*(2), 141–201.

Fan, P., & Gu, W. (2016). Prosodic cues in polite and rude mandarin speech. *2016 10th international symposium on Chinese spoken language processing (ISCSLP)*, 1–4.

Fujisaki, H., Wang, C., Ohno, S., & Gu, W. (2005). Analysis and synthesis of fundamental frequency contours of standard chinese using the command–response model. *Speech communication*, *47*(1-2), 59–70.

Gårding, E. (1979). Sentence intonation in swedish. *Phonetica*, *36*(3), 207–215.

Gårding, E. (1987). Speech act and tonal pattern in standard chinese: Constancy and variation. *Phonetica*, *44*(1), 13–29.

Gårding, E. (1993). On parameters and principles in intonation analysis. *Working papers/Lund University, Department of Linguistics and Phonetics*, *40*, 25–47.

Gussenhoven, C. (2004). The phonology of tone and intonation.

Ho, A. T. (1976). Mandarin tones in relation to sentence intonation and grammatical structure. *Journal of Chinese Linguistics*, 1–13.

Huang, Y. (2017). *The oxford handbook of pragmatics*. Oxford University Press.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Lee, O. J. (2005). *The prosody of questions in beijing mandarin*. The Ohio State University.

Li, C. N., & Thompson, S. A. (1989). *Mandarin chinese: A functional reference grammar*. University of California Press.

Liang, J., & van Heuven, V. J. (2007). Chinese tone and intonation perceived by l1 and l2 listeners. *Tones and tunes*, *2*, 27–61.

Lieberman, P., & Blumstein, S. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge University Press.

Mai, A., Riès, S., Ben-Haim, S., Shih, J. J., & Gentner, T. Q. (2024). Acoustic and language-specific sources for phonemic abstraction from speech. *Nature Communications*, *15*(1), 677.

Ni, J., & Hirose, K. (2006). Quantitative and structural modeling of voice fundamental frequency contours of speech in mandarin. *Speech communication*, *48*(8), 989–1008.

Pierrehumbert, J. (1980). The phonology and phonetics of english intonation. phd diss. *Massachusetts Institute of Technology*.

Prom-On, S., Xu, Y., & Thipakorn, B. (2009). Modeling tone and intonation in mandarin and english as a process of target approximation. *The journal of the Acoustical Society of America*, *125*(1), 405–424.

Romero-Trillo, J., & Sadeghi, Y. (2023). Prosodic pragmatics in context. In J. Romero-Trillo (Ed.), *The cambridge handbook of language in context* (pp. 314–328). Cambridge University Press.

Rumjancev, M. K. (1978). Ton i intonacija v sovremennom kitajskom jazyke(tone and intonation in modern chinese) [As titled and reviewed in Journal of Chinese Linguistics 6, 1978, by A. V. Lyovin, 120–168].

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge university press.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Shen, X.-n. S. (1990). *The prosody of mandarin chinese* (Vol. 118). University of California Press.

Wang, R., Wang, M., & Georgiev, G. V. (2023). Intonation processing of interrogative words in mandarin: An event-related potential study. *Frontiers in Human Neuroscience*, *17*, 1326602.

White, C. M. (1980). Mandarin tone and english intonation: A contrastive analysis.

Xu, Y. (2004). Transmitting tone and intonation simultaneously-the parallel encoding and target approximation (penta) model. *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, 215–220.

Yang, C., & Chan, M. (2010). The perception of mandarin chinese tones and intonation by american learners. *Journal of Chinese Language Teachers Association*, *45*(1), 7–36.

Yang, C., Chu, J., Chen, S., & Xu, Y. (2021). Effects of segments, intonation and rhythm on the perception of l2 accentedness and comprehensibility. *The Acquisition of Chinese as a Second Language Pronunciation: Segments and Prosody*, 233–255.

Yuan, J. (2004). Perception of mandarin intonation. *2004 international symposium on chinese spoken language processing*, 45–48.

Yuan, J., & Shih, C. (2004). Confusability of chinese intonation. *Proceedings of Speech Prosody*, 131–134.

Yuan, J., Shih, C., & Kochanski, G. (2002). Comparison of declarative and interrogative intonation in chinese.

Zhu, Y. (2013). Which is the best listener group?: Perception of chinese emotional prosody by chinese natives, naive dutch listeners and dutch l2 learners of chinese. *Dutch Journal of Applied Linguistics*, *2*(2), 170–183.

Zsiga, E. C. (2024). *The sounds of language: An introduction to phonetics and phonology*. John Wiley & Sons.